

Requestor's Name: T. Wessendorf Serial Number: 09/049,847
Date: 7/21/99 Phone: 8-3967 Art Unit: 1818
SB17

Search Topic:
Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

Please search the structure a-d
with the seq. IDs. 1-4.

also, inventor search.

Thank u.

3X 308-3967.

DO NOT LEAVE THIS SEARCH IN THE CASE!!

Contains confidential **PENDING**
applications data.

RECEIVED
JUL 21 1999
BIO-TECH/CHEM. DIVISION
(STIC)
b-2

STAFF USE ONLY

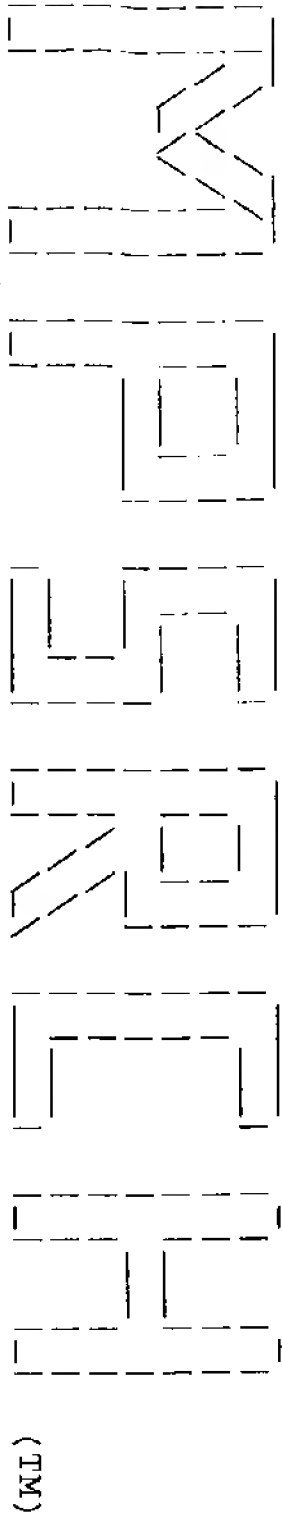
Date completed: 8/5/99
Searcher: _____
Terminal time: 40' + 20' = 60
Elapsed time: _____
CPU time: _____
Total time: 40' / 100
Number of Searches: 2
Number of Databases: 9

Search Site
☐ STIC
☒ CM-1
☐ Pre-S

Type of Search
☐ N.A. Sequence
☒ A.A. Sequence
☒ Structure
☒ Bibliographic

Vendors
☒ IG Suite
☒ STN
☐ Dialog
☐ APS
☐ Geninfo
☐ SDC
☐ DARC/Questel
☐ Other

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:45:44 1999; Maspar time 6.01 Seconds

123.032 Million cell updates/sec

Tabular output not generated.

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFTGTEEL 15

Scoring table: PAM 150
Gap 15

Searched: 403756 seqs, 49297578 residues

Post-processing: Minimum Match 100%
Listing first 1000 summaries
Maximum DB seq length 50

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEW08 20:NEW09

Statistics: Mean 19.535; Variance 56.889; scale 0.343

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	104	100.0	15 12	US-08-716-	Sequence 4, Applicatio	2.55e-04
2	104	100.0	15 15	US-09-049-	Sequence 1, Applicatio	2.55e-04
3	104	100.0	15 15	US-09-046-	Sequence 2, Applicatio	2.55e-04
4	104	100.0	15 8	US-08-406-	Sequence 2, Applicatio	2.55e-04
5	104	100.0	15 1	PCT-US99-0	Sequence 2, Applicatio	2.55e-04
6	104	100.0	15 5	US-08-161-	Sequence 69, Applicati	2.55e-04
7	104	100.0	15 5	US-08-161-	Sequence 69, Applicati	2.55e-04
8	104	100.0	16 1	PCT-US99-0	Sequence 55, Applicati	2.55e-04
9	104	100.0	16 17	US-09-248-	Sequence 55, Applicati	2.55e-04
10	104	100.0	16 10	US-08-577-	Sequence 1, Applicatio	2.55e-04
11	104	100.0	17 4	US-08-057-	Sequence 4, Applicatio	2.55e-04
12	104	100.0	17 6	US-08-218-	Sequence 7, Applicatio	2.55e-04
13	104	100.0	17 4	US-08-060-	Sequence 7, Applicatio	2.55e-04
14	104	100.0	17 7	US-08-328-	Sequence 7, Applicatio	2.55e-04
15	104	100.0	17 9	US-08-488-	Sequence 4, Applicatio	2.55e-04
16	104	100.0	17 9	US-08-472-	Sequence 23, Applicati	2.55e-04
17	104	100.0	17 6	US-08-229-	Sequence 4, Applicatio	2.55e-04
18	104	100.0	17 6	US-08-245-	Sequence 12, Applicati	2.55e-04
19	104	100.0	17 12	US-08-718-	Sequence 7, Applicatio	2.55e-04
20	104	100.0	17 1	PCT-US94-0	Sequence 4, Applicatio	2.55e-04

21	104	100.0	17 9	US-08-472-	Sequence 23, Applicati	2.55e-04
22	104	100.0	17 16	US-09-100-	Sequence 40, Applicati	2.55e-04
23	104	100.0	17 16	US-09-100-	Sequence 32, Applicati	2.55e-04
24	104	100.0	17 14	US-08-926-	Sequence 7, Applicatio	2.55e-04
25	104	100.0	27 4	US-08-057-	Sequence 13, Applicati	2.55e-04
26	104	100.0	27 7	US-08-328-	Sequence 18, Applicati	2.55e-04
27	104	100.0	27 9	PCT-US94-0	Sequence 13, Applicati	2.55e-04
28	104	100.0	27 9	US-08-488-	Sequence 13, Applicati	2.55e-04
29	104	100.0	27 12	US-08-718-	Sequence 18, Applicati	2.55e-04
30	104	100.0	27 6	US-08-218-	Sequence 18, Applicati	2.55e-04
31	104	100.0	27 6	US-08-229-	Sequence 13, Applicati	2.55e-04
32	104	100.0	27 4	US-08-060-	Sequence 18, Applicati	2.55e-04
33	104	100.0	29 7	US-08-328-	Sequence 37, Applicati	2.55e-04
34	104	100.0	29 12	US-08-718-	Sequence 37, Applicati	2.55e-04
35	104	100.0	31 5	US-08-161-	Sequence 63, Applicati	2.55e-04
36	104	100.0	31 5	US-08-161-	Sequence 63, Applicati	2.55e-04
37	104	100.0	37 1	PCT-US94-0	Sequence 57, Applicati	2.55e-04
38	104	100.0	37 9	US-08-488-	Sequence 57, Applicati	2.55e-04
39	104	100.0	37 9	US-08-488-	Sequence 63, Applicati	2.55e-04
40	104	100.0	37 9	PCT-US94-0	Sequence 63, Applicati	2.55e-04
41	104	100.0	47 9	US-08-488-	Sequence 35, Applicati	2.55e-04
42	104	100.0	47 1	PCT-US94-0	Sequence 35, Applicati	2.55e-04
43	104	100.0	47 6	US-08-229-	Sequence 35, Applicati	2.55e-04
44	104	100.0	50 14	US-08-945-	Sequence 8, Applicatio	2.55e-04

Note: Post-processor removed 956 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID US-08-716-249-4 STANDARD: PRT: 15 AA.
AC xxxxxx
XX
DI
XX
DE Sequence 4, Application US/08716249
XX
CC Sequence 4, Application US/08716249
CC GENERAL INFORMATION:
CC APPLICANT: Guichard, Gilles
CC APPLICANT: Muller, Sylviane
CC APPLICANT: Briand, Jean-Paul
CC APPLICANT: Regenmortel, Marc
CC TITLE OF INVENTION: Retropeptides, Antibodies Thereto, and
CC TITLE OF INVENTION: Uses Thereof for Vaccination and In Vitro Diagnosis
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Spencer & Frank
CC STREET: 1100 New York Avenue, Suite 300E
CC CITY: Washington, D.C.
CC COUNTRY: USA
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/716,249
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: WO PCT/FR95/00292
CC FILING DATE: 13-MAR-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Calvetti, Frederick F.
CC REGISTRATION NUMBER: 28,557
CC REFERENCE/DOCKET NUMBER: GROFO 7001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202)414-4000
CC TELEFAX: (202)414-4040

```
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 12; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 2
ID US-09-049-847-1 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
DT
DT
DT
DE Sequence 1, Application US/09049847
XX
CC Sequence 1, Application US/09049847
CC GENERAL INFORMATION:
CC APPLICANT: Bay, Sylvie
CC APPLICANT: Cantacuzene, Daniele
CC APPLICANT: Leclerc, Claude
CC APPLICANT: Lo-Man, Richard
CC TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine
CC TITLE OF INVENTION: comprising the same and use thereof
CC FILE REFERENCE: 102.166A
CC CURRENT APPLICATION NUMBER: US/09/049,847
CC CURRENT FILING DATE: 1998-03-27
CC EARLIER APPLICATION NUMBER: 60/041,726
CC EARLIER FILING DATE: 1997-03-27
CC NUMBER OF SEQ ID NOS: 4
CC SOFTWARE: Patentln Ver. 2.0
CC SEQ ID NO 1
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 3
ID US-09-046-373-2 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
DT
DT
DT
DE Sequence 2, Application US/09046373
XX
CC Sequence 2, Application US/09046373
CC GENERAL INFORMATION:
CC APPLICANT: Sudhir Paul
CC APPLICANT: Larry J. Smith
CC APPLICANT: Gennady Gololobov
CC TITLE OF INVENTION: Methods for identifying inducers and
```

```
CC TITLE OF INVENTION: Inhibitors of Catalytic Antibodies, Compositions and their
CC TITLE OF INVENTION: Use
CC FILE REFERENCE: UNMC 63123
CC CURRENT APPLICATION NUMBER: US/09/046,373
CC CURRENT FILING DATE: 1998-03-23
CC NUMBER OF SEQ ID NOS: 10
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 2
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 15; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 4
ID US-08-406-916B-2 STANDARD; PRT; 15 AA.
XX
AC xxxxxx
DT
DT
DT
DE Sequence 2, Application US/08406916B
XX
CC Sequence 2, Application US/08406916B
CC GENERAL INFORMATION:
CC APPLICANT: STEVENS, VERNON C.
CC TITLE OF INVENTION: VACCINES AND ANTIGENIC CONJUGATES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MILLARD, SIDNEY W.
CC STREET: 7632 SLATE RIDGE BOULEVARD
CC CITY: REYNOLDSBURG
CC STATE: OHIO
CC COUNTRY: USA
CC ZIP: 43068
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
CC COMPUTER: IBM PS/2 COMPATIBLE 486 DX 33
CC OPERATING SYSTEM: MS DOS 6.2
CC SOFTWARE: WORDPERFECT FOR MS DOS 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/406,916B
CC FILING DATE: 27 MAR 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US92/08370
CC FILING DATE: 30 SEP 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KREMBLAS, FRANCIS T., JR.
CC REGISTRATION NUMBER: 22,773
CC REFERENCE/DOCKET NUMBER: URF 2 056 3 3 3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (614) 575 2100
CC TELEFAX: (614) 575 2149
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE:
CC DESCRIPTION: peptide
CC HYPOTHEICAL: NO
CC FRAGMENT TYPE: 830-844 fragment of tetanus toxoid
CC FEATURE:
CC OTHER INFORMATION: T cell epitope
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;
```


Query Match	100.0%;	Score 104;	DB 8;	Length 157;
Best Local Similarity	100.0%;	Pred. No. 2.55e-04;		
Matches	15;	Conservative	0;	Mismatches 0;
				Indels

```
Db      1 QYIKANSKFIGITEL 15
        | | | | | | | | | |
QY      1 QYIKANSKFIGITEL 15
```

RESULT	5	STANDARD;	PRT;	15 AA
ID	PCT-US99-06325-2			

AC	xxxxxxx
XX	
DT	
XX	
DE	Sequence 2, Application PC/TUS9906325

```
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 2
CC LENGTH: 15
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;
```

```

Db 1 QYIKANSKFIGITEL 15
    | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

```

RESULT	6	STANDARD;	PRT;	15 AA.
ID	US-08-161-889-69			

AC	xxxxxx
XX	
DT	
XX	
DE	Sequence 59, Application US/08161889

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 69:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match	100.0%;	Score 104;	DB 5;	Length 15;
Best Local Similarity	100.0%;	Pred. No. 2.55e-04;		
Matches	15; Conservative	0; Mismatches	0; Indels	0; Gaps

```

Db      1 QYIKANSKFIGITEL 15
        |||||||
QY      1 QYIKANSKFIGITEL 15

```

RESULT	7		
ID	US-08-161-889A-69	STANDARD;	PRT; 15 AA

AC	xxxxxxx
XX	
DT	
XX	
DE	Sequence 69, Application US/08161889A

CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889A
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002

CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 69:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
|||
QY 1 QYIKANSKFIGITEL 15

RESULT 8
ID PCT-US99-03055-55 STANDARD; PRT; 16 AA.
XX

AC xxxxxx
DT
XX
DE Sequence 55, Application PC/TUS9903055
XX
CC Sequence 55, Application PC/TUS9903055
CC GENERAL INFORMATION:
CC APPLICANT: Birkett, Ashley J.
CC TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and
CC TITLE OF INVENTION: their Derivatives
CC FILE REFERENCE: SYN-101 4564/69529
CC CURRENT APPLICATION NUMBER: PCT/US99/03055
CC CURRENT FILING DATE: 1999-02-11
CC EARLIER APPLICATION NUMBER: 60/074537
CC EARLIER FILING DATE: 1998-02-12
CC NUMBER OF SEQ ID NOS: 113
CC SOFTWARE: Patentln Ver. 2.0
DE SEQ ID NO 55
CC LENGTH: 16
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC PUBLICATION INFORMATION:
CC JOURNAL: Vaccine
CC VOLUME: 15
CC ISSUE: 4
CC PAGES: 377-
CC DATE: 1997
CC SEQUENCE 16 AA; 1828 MW; 1351 CN;

Query Match 100.0%; Score 104; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
|||
QY 1 QYIKANSKFIGITEL 15

RESULT 9
ID US-09-248-588-55 STANDARD; PRT; 16 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 55, Application US/09248588
XX
CC Sequence 55, Application US/09248588

CC GENERAL INFORMATION:
CC APPLICANT: Birkett, Ashley J.
CC TITLE OF INVENTION: Strategically Modified Hepatitis B Core Proteins and
CC TITLE OF INVENTION: their Derivatives
CC FILE REFERENCE: SYN-101 4564/69529
CC CURRENT APPLICATION NUMBER: US/09/248,588
CC CURRENT FILING DATE: 1999-02-11
CC EARLIER APPLICATION NUMBER: 60/074537
CC EARLIER FILING DATE: 1998-02-12
CC NUMBER OF SEQ ID NOS: 113
CC SOFTWARE: Patentln Ver. 2.0
DE SEQ ID NO 55
CC LENGTH: 16
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC PUBLICATION INFORMATION:
CC JOURNAL: Vaccine
CC VOLUME: 15
CC ISSUE: 4
CC PAGES: 377-
CC DATE: 1997
CC SEQUENCE 16 AA; 1828 MW; 1351 CN;

Query Match 100.0%; Score 104; DB 17; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
|||
QY 1 QYIKANSKFIGITEL 15

RESULT 10
ID US-08-577-106-1 STANDARD; PRT; 16 AA.
XX
AC xxxxxx
DT
XX

DE Sequence 1, Application US/08577106
XX
CC Sequence 1, Application US/08577106
CC GENERAL INFORMATION:
CC APPLICANT: HANSEN, Hans J.
CC TITLE OF INVENTION: USE OF IMMUNOCONJUGATES TO ENHANCE THE
CC TITLE OF INVENTION: EFFICACY OF MULTI-STAGE CASCADE BOOSTING VACCINES
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 3000 K Street, N.W., Suite 500
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20007-5109
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/577,106
CC FILING DATE: 22-DEC-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/268,129
CC FILING DATE: 06-JUL-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: SAXE, Bernhard D.
CC REGISTRATION NUMBER: 28,665
CC REFERENCE/DOCKET NUMBER: 18733/643
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (202)672-5300
CC TELEFAX: (202)672-5399

CC	TELEX:	904136
CC	INFORMATION FOR SEQ ID NO:	1:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	16 amino acids
CC	TYPE:	amino acid
CC	STRANDEDNESS:	single
CC	TOPOLOGY:	linear
SQ	SEQUENCE	16 AA; 1828 MW; 1432 CN;
Query Match 100.0%; Score 104; DB 10; Length 16;		
Best Local Similarity 100.0%; Pred. No. 2.55e-04;		
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	2 QYIKANSKFITEL	16
QY	1 QYIKANSKFITEL	15
RESULT	11	
ID	US-08-057-166-4	STANDARD; PRT; 17 AA.
XX	AC	xxxxxx
DT		
XX		
DE	Sequence 4, Application US/08057166	
XX		
CC	Sequence 4, Application US/08057166	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Ladd, Anna
CC	TITLE OF INVENTION:	Immunogenic Peptides which Contain LHRH
CC	TITLE OF INVENTION:	and a Helper T Cell Epitope for Treatment of Prostate
CC	TITLE OF INVENTION:	Cancer and Induction of Infertility
CC	NUMBER OF SEQUENCES:	17
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	M. Lisa Wilson, United Biomedical Inc.
CC	STREET:	25 Davids Dr.
CC	CITY:	Hauptauge
CC	STATE:	NY
CC	COUNTRY:	U.S.A.
CC	ZIP:	11788
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	Patentin Release #1.0, Version #1.25
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/057,166
CC	FILING DATE:	19930427
CC	CLASSIFICATION:	424
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	Wilson, M. Lisa
CC	REGISTRATION NUMBER:	34,045
CC	REFERENCE/DOCKET NUMBER:	2003
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(516)273-2828
CC	TELEFAX:	(516)273-1717
CC	INFORMATION FOR SEQ ID NO:	4:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	17 amino acids
CC	TYPE:	AMINO ACID
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	peptide
SQ	SEQUENCE	17 AA; 1981 MW; 1619 CN;
Query Match 100.0%; Score 104; DB 4; Length 17;		
Best Local Similarity 100.0%; Pred. No. 2.55e-04;		
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
Db	3 QYIKANSKFITEL	17
QY	1 QYIKANSKFITEL	15

DE	XX	Sequence 7, Application US/08218461
DT	XX	
AC	xxxxxx	
ID	US-08-218-461-7	STANDARD; PRT; 17 AA.
CC	Sequence 7, Application US/08218461	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Chang Yi Wang	
CC	TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED IMMUNOGENS	
CC	TITLE OF INVENTION: FOR THE TREATMENT OF ALLERGY	
CC	NUMBER OF SEQUENCES: 24	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: MORGAN & FINNEGAN	
CC	STREET: 345 PARK AVENUE	
CC	CITY: NEW YORK	
CC	STATE: NEW YORK	
CC	COUNTRY: U.S.A.	
CC	ZIP: 10154	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: FLOPPY DISC	
CC	COMPUTER: IBM PC COMPATIBLE	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: Wordperfect 5.1	
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER: US/08/218,461	
CC	FILING DATE: 10-MAY-1993	
CC	CLASSIFICATION: 514	
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: 07\847,745	
CC	FILING DATE: 06-MAR-1992	
CC	APPLICATION NUMBER: 07\637,364	
CC	FILING DATE: 04-JAN-1991	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Maria C. H. Lin	
CC	REGISTRATION NUMBER: 29,323	
CC	REFERENCE/DOCKET NUMBER: 1151-4061US2	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (212)758-4800	
CC	TELEFAX: (212)751-6849	
CC	TELEX: 421792	
CC	INFORMATION FOR SEQ ID NO: 7:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 17	
CC	TYPE: amino acid	
CC	STRANDEDNESS: not applicable	
CC	TOPOLOGY: unknown	
CC	MOLECULE TYPE: peptide	
CC	SEQUENCE 17 AA; 1981 MW; 1619 CN;	
DB	3 QYIKANSKFIGITEL 17	
OY	1 QYIKANSKFIGITEL 15	
RESULT	13	
ID	US-08-060-798A-7	STANDARD; PRT; 17 AA.
AC	xxxxxx	
DI		
XX		
DE	Sequence 7, Application US/08060798A	

CC Sequence 7, Application US/08060798A
CC GENERAL INFORMATION:
CC APPLICANT: Chang Yi Wang
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISC
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/060,798A
CC FILING DATE: 19930510
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07\847,745
CC FILING DATE: 06-MAR-1992
CC APPLICATION NUMBER: 07\637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4061US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)753-4800
CC TELEFAX: (212)751-5849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17
CC TYPE: amino acid
CC STRANDEDNESS: not applicable
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA; 1981 MW; 1619 CN;
SQ

Query Match 100.0%; Score 104; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 14
ID US-08-328-912B-7 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 7, Application US/08328912B
XX
CC Sequence 7, Application US/08328912B
CC GENERAL INFORMATION:
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 61
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY

CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/328,912B
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17
CC TYPE: amino acid
CC STRANDEDNESS: not applicable
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA; 1981 MW; 1619 CN;
SQ

Query Match 100.0%; Score 104; DB 7; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 15
ID US-08-488-320A-4 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 4, Application US/08488320A
XX
CC Sequence 4, Application US/08488320A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic Peptides Which Contain EHRH
CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate C
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488,351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446,692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229,275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057,166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA; 1981 MW; 1619 CN;
SQ

Query Match 100.0%; Score 104; DB 9; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 OYIKANSKFIGITEL 17
XX |||||||||||
XX 1 OYIKANSKFIGITEL 15

RESULT 16
ID US-08-472-701-23 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
XX
DT
XX
DE Sequence 23, Application US/08472701
XX
CC Sequence 23, Application US/08472701
CC GENERAL INFORMATION:
CC APPLICANT: Griffin, Ann C.
CC APPLICANT: Hickey, William F.
CC TITLE OF INVENTION: Detection and Treatment Methods for
CC TITLE OF INVENTION: Type I Diabetes
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII Text
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/472,701

CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/272,220
CC FILING DATE: 08-JULY-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Decontl, Giulio A., Jr.
CC REGISTRATION NUMBER: 31,503
CC REFERENCE/DOCKET NUMBER: DCI-092DV
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941
CC INFORMATION FOR SEQ ID NO: 23:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC FRAGMENT TYPE: internal
CC SEQUENCE 17 AA; 1969 MW; 1620 CN;
SQ

Query Match 100.0%; Score 104; DB 9; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 OYIKANSKFIGITEL 17
XX |||||||||||
XX 1 OYIKANSKFIGITEL 15

RESULT 17
ID US-08-229-275-4 STANDARD; PRT; 17 AA.
XX
XX
AC xxxxxx
XX
DI
XX
DE Sequence 4, Application US/08229275
XX
CC Sequence 4, Application US/08229275
CC GENERAL INFORMATION:
CC APPLICANT: Iadd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs as
CC TITLE OF INVENTION: vaccines for treatment of prostate cancer and induction
CC TITLE OF INVENTION: infertility
CC NUMBER OF SEQUENCES: 53
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: M. Lisa Wilson
CC STREET: 25 Davids Drive
CC CITY: Hauppauge
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11788
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/229,275
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, M L
CC REGISTRATION NUMBER: 34,045
CC REFERENCE/DOCKET NUMBER: 2003Z
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)273-2828
CC TELEFAX: (516)273-1717
CC INFORMATION FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 18
ID US-08-245-507-12 STANDARD; PRT; 17 AA.
AC xxxxxx
XX
XX
DT
XX
DE Sequence 12, Application US/08245507

CC Sequence 12, Application US/08245507
CC GENERAL INFORMATION:
CC APPLICANT: Houston, Michael
CC APPLICANT: Zhou, Nian
CC APPLICANT: Kay, Cyril
CC APPLICANT: Hodges, Robert
CC APPLICANT: Cachia, Paul
CC APPLICANT: Irvin, Randall
CC TITLE OF INVENTION: Heterodimer Polypeptide Immunogen Carrier
CC TITLE OF INVENTION: Composition and Method
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Dehlinger & Associates
CC STREET: 350 Cambridge Avenue, Suite 250
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94306
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/245,507
CC FILING DATE:
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fabian, Gary R.
CC REGISTRATION NUMBER: 33,875
CC REFERENCE/DOCKET NUMBER: 8900-0009
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 324-0880
CC TELEFAX: (415) 324-0960
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: T antigen, TT2 peptide
SQ SEQUENCE 17 AA; 1981 MW; 1667 CN;

Query Match 100.0%; Score 104; DB 6; Length 17;

Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 19
ID US-08-718-490A-7 STANDARD; PRT; 17 AA.
AC xxxxxx
XX
XX
DT
XX
DE Sequence 7, Application US/08718490A

CC Sequence 7, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang YI
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17
CC TYPE: amino acid
CC STRANDEDNESS: not applicable
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 12; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
1 QYIKANSKFIGITEL 15

RESULT 20
PCT-US94-04832A-4 STANDARD; PRT; 17 AA.

AC xxxxxx

DE Sequence 4, Application PC/TUS9404832A

CC Sequence 4, Application PC/TUS9404832A

CC GENERAL INFORMATION:

CC APPLICANT: Ladd, Anna

CC APPLICANT: Wang, Chang Yi

CC APPLICANT: Lamb, Timothy

CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs

CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines

CC NUMBER OF SEQUENCES: 114

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE:

CC STREET: 400 Garden City Plaza

CC CITY: Garden City

CC STATE: NY

CC COUNTRY: US

CC ZIP: 11530

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US94/04832A

CC FILING DATE: 13-APR-1994

CC CLASSIFICATION:

CC ATTORNEY/AGENT INFORMATION:

CC NAME:

CC REGISTRATION NUMBER:

CC REFERENCE/DOCKET NUMBER:

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (516)742-4343

CC TELEFAX: (516)742-4366

CC INFORMATION FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 17 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 1; Length 17;

Best Local Similarity 100.0%; Pred. No. 2.55e-04;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17

1 QYIKANSKFIGITEL 15

RESULT 21
US-08-472-704-23 STANDARD; PRT; 17 AA.

AC xxxxxx

XX

DT

DE Sequence 23, Application US/08472704

CC

CC Sequence 23, Application US/08472704

GENERAL INFORMATION:

CC APPLICANT: Hickey, William F.

CC APPLICANT: Griffin, Ann C.

CC TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting and

CC TITLE OF INVENTION: Treating Type I Diabetes

CC NUMBER OF SEQUENCES: 23

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: LAHIVE & COCKFIELD

CC STREET: 60 State Street, suite 510

CC CITY: Boston

CC STATE: Massachusetts

CC COUNTRY: USA

CC ZIP: 02109-1875

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: ASCII Text

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/472,704

CC FILING DATE:

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/272,220

CC FILING DATE: 08-JULY-1994

CC CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Deconti, Giulio A., Jr.

CC REGISTRATION NUMBER: 31,503

CC REFERENCE/DOCKET NUMBER: DCI-092

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617)227-7400

CC TELEFAX: (617)227-5941

CC INFORMATION FOR SEQ ID NO: 23:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 17 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: internal

CC SEQUENCE 17 AA; 1969 MW; 1620 CN;

Query Match 100.0%; Score 104; DB 9; Length 17;

Best Local Similarity 100.0%; Pred. No. 2.55e-04;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17

1 QYIKANSKFIGITEL 15

RESULT 22
US-09-100-409-40 STANDARD; PRT; 17 AA.

AC xxxxxx

XX

DT

XX

DE Sequence 40, Application US/09100409

CC Sequence 40, Application US/09100409

CC GENERAL INFORMATION:

CC APPLICANT: Wang, Chang Yi

CC TITLE OF INVENTION: PEPTIDE COMPOSITION FOR

CC TITLE OF INVENTION: PREVENTION AND TREATMENT OF HIV INFECTION AND

CC TITLE OF INVENTION: IMMUNE DISORDERS

CC NUMBER OF SEQUENCES: 60

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: MORGAN & FINNEGAN

CC STREET: 345 Park Avenue

CC CITY: New York

CC STATE: NY

CC COUNTRY: USA

CC ZIP: 10154-0054
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/100,409
CC FILING DATE:
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER: 1151-4154
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-751-6849
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 40:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYKANSKFIGITEL 17
OY 1 QYKANSKFIGITEL 15

RESULT 23
ID US-09-100-415-32 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 32, Application US/09100415
XX
CC Sequence 32, Application US/09100415
CC GENERAL INFORMATION:
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: SYNTHETIC SOMATOSTATIN IMMUNOGEN FOR
CC TITLE OF INVENTION: GROWTH PROMOTION IN FARM ANIMALS
CC NUMBER OF SEQUENCES: 45
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Morgan & Finnegan
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154-0054
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/100,415
CC FILING DATE:
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER: 1151-4155
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-751-6849
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 32:
CC

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYKANSKFIGITEL 17
OY 1 QYKANSKFIGITEL 15

RESULT 24
ID US-08-926-296-7 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 7, Application US/08926296
XX
CC Sequence 7, Application US/08926296
CC GENERAL INFORMATION:
CC APPLICANT: Walfield, Alan M.
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: Synthetic Ige Membrane Anchor
CC TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/926,296
CC FILING DATE: 05-SEP-1997
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/328,519
CC FILING DATE: 26-OCT-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lin, Maria C.H.
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4117
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-751-6849
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 14; Length 17;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYKANSKFIGITEL 17
OY 1 QYKANSKFIGITEL 15

QY 1 QYIKANSKFIGITEL 15

RESULT 25

ID US-08-057-166-13 STANDARD; PRT; 27 AA.

AC xxxxxx

DE Sequence 13, Application US/08057166

CC Sequence 13, Application US/08057166

CC GENERAL INFORMATION:

CC APPLICANT: Ladd, Anna

CC APPLICANT: Wang, Chang Yi

CC TITLE OF INVENTION: Immunogenic Peptides which Contain LHRH

CC TITLE OF INVENTION: and a Helper T Cell Epitope for Treatment of Prostate

CC TITLE OF INVENTION: Cancer and Induction of Infertility

CC NUMBER OF SEQUENCES: 17

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: M. Lisa Wilson, United Biomedical Inc.

CC STREET: 25 Davids Dr.

CC CITY: Hauppauge

CC STATE: NY

CC COUNTRY: U.S.A.

CC ZIP: 11788

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentln Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/057,166

CC FILING DATE: 19930427

CC CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Wilson, M. Lisa

CC REGISTRATION NUMBER: 34,045

CC REFERENCE/DOCKET NUMBER: 2003

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (516)273-2828

CC TELEFAX: (516)273-1717

CC INFORMATION FOR SEQ ID NO: 13:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 27 amino acids

CC TYPE: AMINO ACID

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 27 AA; 3165 MW; 4134 CN;

QY Db 3 QYIKANSKFIGITEL 17

Query Match 100.0%; Score 104; DB 4; Length 27;

Best Local Similarity 100.0%; Pred. No. 2.55e-04;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 26

ID US-08-328-912B-18 STANDARD; PRT; 27 AA.

AC xxxxxx

DE Sequence 18, Application US/08328912B

CC Sequence 18, Application US/08328912B

CC GENERAL INFORMATION:

CC APPLICANT: Wang, Chang Yi

CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED

CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY

CC NUMBER OF SEQUENCES: 61

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Maria C.H. Lin

CC STREET: 345 Park Avenue

CC CITY: New York

CC STATE: NY

CC COUNTRY: USA

CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Wordperfect 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/328,912B

CC FILING DATE: 25-OCT-1994

CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/218,461

CC FILING DATE: 28-MAR-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/060,798

CC FILING DATE: 10-MAY-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 07/847,745

CC FILING DATE: 06-MAR-1992

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 07/637,364

CC FILING DATE: 04-JAN-1991

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Maria C.H. Lin

CC REGISTRATION NUMBER: 29,323

CC REFERENCE/DOCKET NUMBER: 1151-4061US4

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 212-758-4800

CC TELEFAX: 212-751-6849

CC INFORMATION FOR SEQ ID NO: 18:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 27

CC TYPE: amino acid

CC STRANDEDNESS: not applicable

CC TOPOLOGY: unknown

CC MOLECULE TYPE: peptide

CC SEQUENCE 27 AA; 3081 MW; 4686 CN;

QY Db 3 QYIKANSKFIGITEL 17

Query Match 100.0%; Score 104; DB 7; Length 27;

Best Local Similarity 100.0%; Pred. No. 2.55e-04;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27

ID PCT-US94-04832A-13 STANDARD; PRT; 27 AA.

AC xxxxxx

DE Sequence 13, Application PC/TUS9404832A

CC Sequence 13, Application PC/TUS9404832A

CC GENERAL INFORMATION:

CC APPLICANT: Ladd, Anna

CC APPLICANT: Wang, Chang Yi

CC APPLICANT: Zamb, Timothy

CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs

CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

CC NUMBER OF SEQUENCES: 114

CC CORRESPONDENCE ADDRESS:

CC ADDRESS: 400 Garden City Plaza
CC STREET: Garden City
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 27 AA; 3165 MW; 4134 CN;
SQ
Query Match 100.0%; Score 104; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15
RESULT 28
ID US-08-488-320A-13 STANDARD; PRT; 27 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 13, Application US/08488320A
XX
CC Sequence 13, Application US/08488320A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic Peptides which Contain LHRH
CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Can
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488,351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446,692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229,275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057,166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 27 AA; 3165 MW; 4134 CN;
SQ
Query Match 100.0%; Score 104; DB 9; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15
RESULT 29
ID US-08-718-490A-18 STANDARD; PRT; 27 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 18, Application US/08718490A
XX
CC Sequence 18, Application US/08718490A
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical, Inc. & WANG, Chang Yi
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 32
ID US-08-060-798A-18 STANDARD; PRT; 27 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX

Sequence 18, Application US/08060798A

CC Sequence 18, Application US/08060798A
CC GENERAL INFORMATION:

CC APPLICANT: Chang Yi Wang
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: FLOPPY DISC
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/060,798A
CC FILING DATE: 19930510
CC CLASSIFICATION: 530

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 07\847,745
CC FILING DATE: 06-MAR-1992
CC APPLICATION NUMBER: 07\637,364
CC FILING DATE: 04-JAN-1991

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Maria C. H. Lin

CC REGISTRATION NUMBER: 29,323

CC REFERENCE/DOCKET NUMBER: 1151-4061US2

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212)758-4800

CC TELEFAX: (212)751-6849

CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 18:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 27

CC TYPE: amino acid

CC STRANDEDNESS: not applicable

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 27 AA; 3081 MW; 4686 CN;

Query Match 100.0%; Score 104; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 33
ID US-08-328-912B-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX

DT
XX
DE
XX
Sequence 37, Application US/08328912B

CC Sequence 37, Application US/08328912B
CC GENERAL INFORMATION:

CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 61
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/328,912B
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION: 424

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Maria C.H. Lin

CC REGISTRATION NUMBER: 29,323

CC REFERENCE/DOCKET NUMBER: 1151-4061US4

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 212-751-6849

CC INFORMATION FOR SEQ ID NO: 37:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 29 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 29 AA; 3195 MW; 5252 CN;

Query Match 100.0%; Score 104; DB 7; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 34
ID US-08-718-490A-37 STANDARD; PRT; 29 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 37, Application US/08718490A
Sequence 37, Application US/08718490A
GENERAL INFORMATION:

CC APPLICANT: United Biomedical, Inc. & WANG, Chang yi
CC TITLE OF INVENTION: SYNTHETIC PEPTIDE BASED
CC TITLE OF INVENTION: IMMUNOGENS FOR THE TREATMENT OF ALLERGY
CC NUMBER OF SEQUENCES: 62
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/718,490A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/328,912
CC FILING DATE: 25-OCT-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/218,461
CC FILING DATE: 28-MAR-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/060,798
CC FILING DATE: 10-MAY-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/847,745
CC FILING DATE: 06-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/637,364
CC FILING DATE: 04-JAN-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4061US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC INFORMATION FOR SEQ ID NO: 37:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 29 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 29 AA; 3195 MW; 5252 CN;
SQ

Query Match 100.0%; Score 104; DB 12; Length 29;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
OY 1 QYIKANSKFIGITEL 15

RESULT 35
ID US-08-161-889-63 STANDARD; PRI; 31 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 63, Application US/08161889
XX
CC Sequence 63, Application US/08161889
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes

CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3583 MW; 5387 CN;
SQ

Query Match 100.0%; Score 104; DB 5; Length 31;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 QYIKANSKFIGITEL 24
OY 1 QYIKANSKFIGITEL 15

RESULT 36
ID US-08-161-889A-63 STANDARD; PRI; 31 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 63, Application US/08161889A
XX
CC Sequence 63, Application US/08161889A
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 74
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/161,889A
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 31 AA; 3583 MW; 5387 CN;

Query Match 100.0%; Score 104; DB 5; length 31;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 10 QYIKANSKFIGITEL 24
QY 1 QYIKANSKFIGITEL 15

RESULT 37
ID PCT-US94-04832A-57 STANDARD; PRT; 37 AA.
XX xxxxxx
AC xxxxxx
DT
DE
XX
XX

Sequence 57, Application PC/TUS9404832A

GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic LHRH peptide constructs
TITLE OF INVENTION: And synthetic universal immune stimulators for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: US
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04832A
FILING DATE: 13-APR-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
SEQUENCE 37 AA; 4060 MW; 7526 CN;

CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 1; length 37;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15

RESULT 38
ID US-08-488-320A-57 STANDARD; PRT; 37 AA.
XX xxxxxx
AC xxxxxx
DT
DE
XX
XX

Sequence 57, Application US/08488320A

GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic peptides which contain LHRH
TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate C
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,320A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,351
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/446,692
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/229,275
FILING DATE: 14-APR-1994
APPLICATION NUMBER: 08/057,166
FILING DATE: 27-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146US4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFITTEL 35
QY 1 QYIKANSKFITTEL 15

RESULT 39
ID US-08-488-320A-63 STANDARD; PRT; 37 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 63, Application US/08488320A
XX

Sequence 63, Application US/08488320A
GENERAL INFORMATION:
APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH
TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Cancer
TITLE OF INVENTION: And Induction Of Infertility
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,320A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/488,351
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/446,692
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: 08/229,275
FILING DATE: 14-APR-1994
APPLICATION NUMBER: 08/057,166
FILING DATE: 27-APR-1993

ATTORNEY/AGENT INFORMATION:
NAME: Maria C. H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146US4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)758-4800
TELEFAX: (212)751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 100.0%; Score 104; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFITTEL 19
QY 1 QYIKANSKFITTEL 15

RESULT 40
ID PCT-US94-04832A-63 STANDARD; PRT; 37 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 63, Application PC/TUS9404832A
XX

Sequence 63, Application PC/TUS9404832A
GENERAL INFORMATION:
APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic LHRH peptide constructs
TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:

ADDRESSEE:
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: NY
COUNTRY: US
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04832A
FILING DATE: 13-APR-1994

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME:

REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516)742-4343
TELEFAX: (516)742-4366
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 37 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 100.0%; Score 104; DB 1; Length 37;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFITTEL 19
QY 1 QYIKANSKFITTEL 15

RESULT 41
ID US-08-488-320A-35 STANDARD; PRT; 47 AA.
AC xxxxxx
XX
DT
XX
DE Sequence 35, Application US/08488320A
XX

Sequence 35, Application US/08488320A
GENERAL INFORMATION:

CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH
CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Cancer
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488,351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446,692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229,275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057,166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;
SQ
Query Match 100.0%; Score 104; DB 9; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
RESULT 42
ID PCT-US94-04832A-35 STANDARD; PRT: 47 AA.
XX PCT-US94-04832A-35
AC xxxxxx
XX
DT
XX
DE Sequence 35, Application PC/TUS9404832A
XX
CC Sequence 35, Application PC/TUS9404832A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC

CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;
SQ
Query Match 100.0%; Score 104; DB 1; Length 47;
Best Local Similarity 100.0%; Pred. No. 2.55e-04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
RESULT 43
ID US-08-229-275-35 STANDARD; PRT: 47 AA.
XX US-08-229-275-35
AC xxxxxx
XX
DT
XX
DE Sequence 35, Application US/08229275
XX
CC Sequence 35, Application US/08229275
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs as
CC TITLE OF INVENTION: vaccines for treatment of prostate cancer and induction
CC TITLE OF INVENTION: infertility
CC NUMBER OF SEQUENCES: 53
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: M. Lisa Wilson
CC STREET: 25 Davids Drive
CC CITY: Hauppauge
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11788
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/229,275
CC

CC FILING DATE: 13-APR-1994
 CC CLASSIFICATION: 424
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Wilson, M. L.
 CC REGISTRATION NUMBER: 34,045
 CC REFERENCE/DOCKET NUMBER: 2003Z
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (516)273-2828
 CC TELEFAX: (516)273-1717
 CC INFORMATION FOR SEQ ID NO: 35:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 47 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 47 AA; 5243 MW; 12301 CN;
 SQ
 Query Match 100.0%; Score 104; DB 6; Length 47;
 Best Local Similarity 100.0%; Pred. No. 2.55e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 21 QYIKANSKFIGITEL 35
 QY 1 QYIKANSKFIGITEL 15
 RESULT 44
 ID US-08-945-289-8 STANDARD; PRT; 50 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 8, Application US/08945289
 XX
 CC Sequence 8, Application US/08945289
 CC GENERAL INFORMATION:
 CC APPLICANT: Rittershaus, Charles, W.
 CC APPLICANT: Thomas, Lawrence J.
 CC TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
 CC TITLE OF INVENTION: TRANSFER PROTEIN (CETP) ACTIVITY
 CC NUMBER OF SEQUENCES: 9
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Yankwich & Associates
 CC STREET: 130 Bishop Allen Drive
 CC CITY: Cambridge
 CC STATE: Massachusetts
 CC COUNTRY: USA
 CC ZIP: 02139
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: Windows 95
 CC SOFTWARE: Word 97
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/945,289
 CC FILING DATE: October 17, 1997
 CC CLASSIFICATION: 514
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 08/432,483
 CC FILING DATE: May 1, 1995
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Leon R. Yankwich
 CC REGISTRATION NUMBER: 30,237
 CC REFERENCE/DOCKET NUMBER: TCS-411.1P US
 CC INFORMATION FOR SEQ ID NO: 8:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 50 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL:
 CC ANTI-SENSE:

CC FEATURE:
 CC NAME/KEY:
 CC LOCATION:
 SQ SEQUENCE 50 AA; 5821 MW; 14126 CN;
 Query Match 100.0%; Score 104; DB 14; Length 50;
 Best Local Similarity 100.0%; Pred. No. 2.55e-04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 2 QYIKANSKFIGITEL 16
 QY 1 QYIKANSKFIGITEL 15
 Search completed: Wed Aug 4 15:47:54 1999
 Job time : 70 secs.

This Page Blank (uspto)

SQ SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 3; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
1 QYIKANSKFIGITEL 15

RESULT 2
ID US-08-460-502-7 STANDARD; PRT; 15 AA.

AC xxxxxx

DE Sequence 7, Application US/08460502

CC Sequence 7, Application US/08460502
CC Patent No. 5843464

CC GENERAL INFORMATION:

CC APPLICANT: Bakaletz, Lauren O.

CC APPLICANT: Kaumaya, Parvin T.

CC TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides

CC NUMBER OF SEQUENCES: 11

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSEE: Calfee, Halter and Griswold

CC STREET: 800 Superior Avenue

CC CITY: Cleveland

CC STATE: Ohio

CC COUNTRY: U.S.A.

CC ZIP: 44114-2688

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/460,502

CC FILING DATE:

CC CLASSIFICATION: 424

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Goltick, Mary E.

CC REGISTRATION NUMBER: 34,829

CC REFERENCE/DOCKET NUMBER: 22727/00120

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (216) 622-8458

CC TELEFAX: (216) 241-0816

CC INFORMATION FOR SEQ ID NO: 7:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 15 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
1 QYIKANSKFIGITEL 15

RESULT 3
ID US-08-661-052-6 STANDARD; PRT; 15 AA.

AC xxxxxx
XX

DT Sequence 6, Application US/08661052

DE Sequence 6, Application US/08661052
XX Patent No. 5837243

CC GENERAL INFORMATION:

CC APPLICANT: Yashwant M. Deo

CC APPLICANT: Joel Goldstein

CC APPLICANT: Robert Graziano

CC APPLICANT: Chezia Somasundaram

CC TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED

CC TITLE OF INVENTION: OF ANTI-FC RECEPTOR ANTIBODIES

CC NUMBER OF SEQUENCES: 16

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: LAHIVE & COCKFIELD

CC STREET: 60 State Street, Suite 510

CC CITY: Boston

CC STATE: Massachusetts

CC COUNTRY: USA

CC ZIP: 02109-1875

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/661,052

CC FILING DATE:

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/484,172

CC FILING DATE: 07-JUNE-1995

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Arnold, Beth E.

CC REGISTRATION NUMBER: 35,430

CC REFERENCE/DOCKET NUMBER: MXI-043CP

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617)227-7400

CC TELEFAX: (617)227-5941

CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 15 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC FRAGMENT TYPE: Internal

CC SEQUENCE 15 AA; 1725 MW; 1271 CN;

Query Match 100.0%; Score 104; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
1 QYIKANSKFIGITEL 15

RESULT 4
ID US-08-319-704-10 STANDARD; PRT; 15 AA.

AC xxxxxx

DE Sequence 10, Application US/08319704

CC Sequence 10, Application US/08319704
CC Patent No. 5814617

CC GENERAL INFORMATION:

CC APPLICANT: Hoffman, Stephen L.

CC APPLICANT: Charoenvit, Yupin

CC APPLICANT: Hedstrom, Richard C.

CC APPLICANT: Doolan, Denise L.

CC TITLE OF INVENTION: Protective 17 kDa Malaria Hepatic and
CC TITLE OF INVENTION: Erythrocytic Stage Immunogen and Gene
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Naval Medical R & D Command
CC STREET: Bldg 1, T-12, 8901 Wisconsin Avenue
CC CITY: Bethesda
CC STATE: Maryland
CC COUNTRY: U.S.A
CC ZIP: 20889-5606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/319,704
CC FILING DATE: 07-OCT-1994
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: A. David Spevack
CC REGISTRATION NUMBER: 24,743
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (301) 295-6759
CC TELEFAX: (301) 295-1022
CC INFORMATION FOR SEQ ID NO: 10:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 15 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 15 AA; 1725 MW; 1271 CN;
SQ

Query Match 100.0%; Score 104; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 QYIKANSKFIGITEL 15
QY 1 QYIKANSKFIGITEL 15

RESULT 5
ID PCT-US95-13841-7 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 7, Application PC/TUS9513841
XX
CC Sequence 7, Application PC/TUS9513841
CC GENERAL INFORMATION:
CC APPLICANT: United Biomedical Inc; Walfield, Alan M.;
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: Synthetic Ige Membrane Anchor
CC TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
CC NUMBER OF SEQUENCES: 32
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Wordperfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/13841

CC FILING DATE: 25-OCT-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/328,519
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lin, Maria C.H.
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4117
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA; 1981 MW; 1619 CN;
SQ

Query Match 100.0%; Score 104; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 6
ID PCT-US95-08596-23 STANDARD; PRT; 17 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 23, Application PC/TUS9508596
XX
CC Sequence 23, Application PC/TUS9508596
CC GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: Proinsulin Peptide Compounds for Detecting
CC TITLE OF INVENTION: and Treating Type I Diabetes
CC NUMBER OF SEQUENCES: 23
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: LAHIVE & COCKFIELD
CC STREET: 60 State Street, suite 510
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02109-1875
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASCII Text
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/08596
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/272,220
CC FILING DATE: 08-JULY-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: DeConti, Giulio A., Jr.
CC REGISTRATION NUMBER: 31,503
CC REFERENCE/DOCKET NUMBER: DCI-092PC
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)227-7400
CC TELEFAX: (617)227-5941

```

CC      INFORMATION FOR SEQ ID NO: 23:
CC      SEQUENCE CHARACTERISTICS:
CC          LENGTH: 17 amino acids
CC          TYPE: amino acid
CC          TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
CC      FRAGMENT TYPE: internal
CC      SEQUENCE 17 AA; 1969 MW; 1620 CN;

Db      3 QYIKANSKEIGITEL 17
QY      1 QYIKANSKEIGITEL 15

Query Match      100.0%; Score 104; DB 3; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7
ID      US-08-488-351A-4      STANDARD; PRT; 17 AA.
XX      xxxxxx
AC
DT
XX
XX
DE      Sequence 4, Application US/08488351A
XX
CC      Sequence 4, Application US/08488351A
CC      Patent No. 5843446
CC      GENERAL INFORMATION:
CC      APPLICANT: Ladd, Anne.
CC      APPLICANT: Wang, Chang Yi
CC      APPLICANT: Zamb, Timothy
CC      TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC      TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC      NUMBER OF SEQUENCES: 114
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Maria C.H. Lin
CC      STREET: 345 Park Avenue
CC      CITY: New York
CC      STATE: NY
CC      COUNTRY: US
CC      ZIP: 10154-0053
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: patentin Release #1.0, Version #1.25
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/08/488,351A
CC      FILING DATE: 7-JUN-1995
CC      CLASSIFICATION: 424
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/446,692
CC      FILING DATE: 7-JUN-1995
CC      CLASSIFICATION: 424
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/229,275
CC      FILING DATE: 14-APR-1994
CC      CLASSIFICATION: 424
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: US 08/057,166
CC      FILING DATE: 27-APR-1992
CC      CLASSIFICATION: 424
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Maria C.H. Lin
CC      REGISTRATION NUMBER: 29,323
CC      REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (212)415-8745
CC      TELEFAX: (516)751-6849
CC      INFORMATION FOR SEQ ID NO: 4:
CC      SEQUENCE CHARACTERISTICS:

```

```

CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

RESULT 8 STANDARD; PRT; 17 AA.
ID US-08-446-692-4
XX AC xxxxxx
XX DT
XX DE
DE Sequence 4, Application US/08446692
CC Sequence 4, Application US/08446692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vacciness
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 17 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 17 AA; 1981 MW; 1619 CN;

Query Match 100.0%; Score 104; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 QYIKANSKFIGITEL 17
QY 1 QYIKANSKFIGITEL 15

```


RESULT 9
ID US-08-446-692-13 STANDARD; PRT; 27 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 13, Application US/08446692
CC
CC Sequence 13, Application US/08446692
CC Patent No. 5759551
CC
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQ SEQUENCE 27 AA; 3165 MW; 4134 CN;
Query Match 100.0%; Score 104; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFTGTTEL 17
QY 1 QYIKANSKFTGTTEL 15
RESULT 10
ID US-08-488-351A-13 STANDARD; PRT; 27 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 13, Application US/08488351A
CC
CC Sequence 13, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy

CC
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 27 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQ SEQUENCE 27 AA; 3165 MW; 4134 CN;
Query Match 100.0%; Score 104; DB 2; Length 27;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 3 QYIKANSKFTGTTEL 17
QY 1 QYIKANSKFTGTTEL 15
RESULT 11
ID PCT-US93-11703-63 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 63, Application PC/TUS9311703
CC
CC GENERAL INFORMATION:
CC APPLICANT: Chiron Mimotopes Pty. Ltd.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville

CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:
CC PRIORITY APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3583 MW; 5387 CN;
SQ

Query Match 100.0%; Score 104; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 10 QYKANSKFIGITEL 24
QY 1 QYKANSKFIGITEL 15
RESULT 12
ID US-08-446-692-63 STANDARD; PRT; 37 AA.
XX
XX xxxxxx
DE Sequence 63, Application US/08446692
XX
CC Sequence 63, Application US/08446692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995

CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 37 AA; 4060 MW; 8502 CN;
SQ

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 5 QYKANSKFIGITEL 19
QY 1 QYKANSKFIGITEL 15
RESULT 13
ID US-08-488-351A-57 STANDARD; PRT; 37 AA.
XX
XX xxxxxx
DE Sequence 57, Application US/08488351A
XX
CC Sequence 57, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin

CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15

RESULT 14
ID US-08-446-692-57 STANDARD; PRT; 37 AA.
XX xxxxxx
DT
DT
DE Sequence 57, Application US/084446692
XX
CC Sequence 57, Application US/084446692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 57:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 7526 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15

RESULT 15
ID US-08-488-351A-63 STANDARD; PRT; 37 AA.
XX xxxxxx
AC xxxxxx
DT
DT
DE Sequence 63, Application US/08488351A
XX
CC Sequence 63, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentln Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 63:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 37 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 37 AA; 4060 MW; 8502 CN;

Query Match 100.0%; Score 104; DB 2; Length 37;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 5 QYIKANSKFIGITEL 19

```
|||||
QY 1 QYIKANSKFIGITEL 15

RESULT 16
ID US-08-488-351A-35 STANDARD; PRT; 47 AA.
AC xxxxxx
DT
DT
DT
DE Sequence 35, Application US/08488351A
XX
CC Sequence 35, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 100.0%; Score 104; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
```

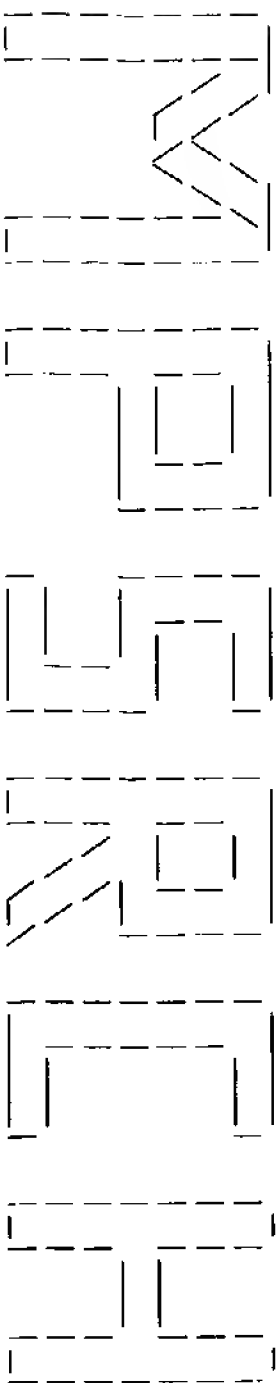
```
|||||
QY 1 QYIKANSKFIGITEL 15

RESULT 17
ID US-08-446-692-35 STANDARD; PRT; 47 AA.
AC xxxxxx
DT
DT
DT
DE Sequence 35, Application US/08446692
XX
CC Sequence 35, Application US/08446692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESSES:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 47 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 47 AA; 5243 MW; 12301 CN;

Query Match 100.0%; Score 104; DB 2; Length 47;
Best Local Similarity 100.0%; Pred. No. 1.64e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 21 QYIKANSKFIGITEL 35
QY 1 QYIKANSKFIGITEL 15
```

Search completed: Wed Aug 4 15:45:41 1999
Job time : 5 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:25:15 1999; MasPar time 4.84 Seconds
173.711 Million cell updates/sec
Tabular output not generated.

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSFWLRYPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 29.518; Variance 48.854; scale 0.604

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	158	100.0	1315	1	BTCLTN	tentoxylysin (EC 3.4. 3.74e-18	
2	95	60.1	1268	2	S33411	botulinum neurotoxin	3.74e-05
3	95	60.1	1297	2	S39791	neurotoxin - Clostrid	3.74e-05
4	93	58.9	366	2	S48110	neurotoxin type F - C	8.82e-05
5	93	58.9	369	2	S48109	neurotoxin type F - C	8.82e-05
6	93	58.9	1274	2	I40813	neurotoxin type F - C	8.82e-05
7	93	58.9	1291	2	I40631	non-protolytic botul	8.82e-05
8	93	58.9	1291	1	A48940	bontoxilysin (EC 3.4. 3.15e-04	
9	90	57.0	1296	1	BTCLAB	bontoxilysin (EC 3.4. 7.30e-04	
10	88	55.7	367	2	S48106	neurotoxin type F - C	7.30e-04
11	88	55.7	1251	2	JH0257	botulinum neurotoxin	7.30e-04
12	88	55.7	1251	2	JH0256	botulinum neurotoxin	7.30e-04
13	88	55.7	1252	2	S21178	botulinum neurotoxin	7.30e-04
14	85	53.8	1296	2	I40645	botulinum neurotoxin	2.53e-03
15	84	53.2	464	1	MNVUWC	nonstructural protein	3.82e-03
16	84	53.2	467	1	MNVUW1	nonstructural protein	3.82e-03
17	81	51.3	1285	2	S70582	botulinum neurotoxin	1.29e-02
18	81	51.3	1291	2	A49777	botulinum neurotoxin	1.29e-02
19	81	51.3	1291	2	S46431	botulinum neurotoxin	1.29e-02
20	77	48.7	449	2	S23158	nucleocapsid protein	6.34e-02
21	76	48.1	1196	2	S46430	botulinum neurotoxin-	9.34e-02
22	76	48.1	1196	2	J01467	toxin, nontoxic compo	9.34e-02
23	72	45.6	406	2	F70930	hypothetical protein	4.34e-01

24	72	45.6	1276	2	S11455	botulinum neurotoxin	4.34e-01
25	70	44.3	1162	2	I40817	botulinum toxin nonto	9.20e-01
26	70	44.3	1162	2	A47708	progenitor toxin nont	9.20e-01
27	68	43.0	699	2	I38073	nucleolar phosphoprot	1.92e+00
28	67	42.4	152	2	D64943	probable membrane pro	2.77e+00
29	67	42.4	789	1	QXBY32	gene cox1 intron 2 pr	2.77e+00
30	67	42.4	854	2	S78644	gene cox1 intron 2 pr	2.77e+00
31	66	41.8	241	2	S17822	spheroidene monooxyge	3.97e+00
32	66	41.8	244	2	H70591	hypothetical protein	3.97e+00
33	66	41.8	591	2	S04401	spheroidene monooxyge	3.97e+00
34	66	41.8	1165	2	I40644	botulinum neurotoxin	3.97e+00
35	66	41.8	1193	2	S68218	botulinum neurotoxin	3.97e+00
36	66	41.8	1193	3	JC4901	nontoxic-nonhemagglu	3.97e+00
37	65	41.1	203	2	A36886	surface protein pag n	5.67e+00
38	65	41.1	836	2	F17546	probable DNA gyrase c	5.67e+00
39	65	41.1	854	2	S17995	gene COX1 intron 1 pr	5.67e+00
40	63	39.9	278	2	T00613	hypothetical protein	1.15e+01
41	63	39.9	547	2	S51475	cytochrome P450 - mal	1.15e+01
42	63	39.9	605	1	ABPGS	serum albumin precurs	1.15e+01
43	63	39.9	728	2	S57142	hypothetical protein	1.15e+01
44	63	39.9	3972	2	S75251	hypothetical protein	1.15e+01
45	62	39.2	380	2	S33720	NS1 protein - dengue	1.62e+01

ALIGNMENTS

RESULT ENTRY	1	ALIGNMENTS
BTCLTN	#type complete	
tentoxylysin (EC 3.4.24.68)	precursor - Clostridium tetani	
tetanus neurotoxin		
ORGANISM	#formal_name Clostridium tetani	
DATE	31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 26-Feb-1999	
ACCESSIONS	A25689; A25757; A25194; B25194; A60759; S69348; S09364	
REFERENCE	A25689	
#authors	Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.; Habermann, E.; Niemann, H.	
#journal	EMBO J. (1986) 5:2495-2502	
#title	Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins.	
#cross-references	MUID:87053814	
#accession	A25689	
##molecule_type	DNA	
##residues	1-1315 #label EIS	
##cross-references	GB:X04436; NID:g40769; PID:g40770	
REFERENCE	A25757	
#authors	Fairweather, N.F.; Lyness, V.A.	
#journal	Nucleic Acids Res. (1986) 14:7809-7812	
#title	The complete nucleotide sequence of tetanus toxin.	
#cross-references	MUID:87040747	
#accession	A25757	
##molecule_type	DNA	
##residues	1-1315 #label FAI	
##cross-references	GB:X06214; NID:g40773; PID:g40774	
##experimental_source	strain CN3911	
REFERENCE	A25194	
#authors	Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.	
#journal	J. Bacteriol. (1986) 165:21-27	
#title	Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.	
#cross-references	MUID:86085672	
#accession	A25194	
##molecule_type	DNA	
##residues	743-1315 #label FA2	
##cross-references	GB:M12739; NID:g144920; PID:g144921	
#accession	B25194	
##molecule_type	protein	
##residues	865-894 #label FA3	
REFERENCE	A60759	
#authors	Matsuda, M.; Iel, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.	
#journal	Infect. Immun. (1989) 57:3588-3593	
#title	Isolation, purification, and characterization of fragment B,	

the NH-2-terminal half of the heavy chain of tetanus toxin.

#cross-references MUID:90035436
#accession A60759
##molecule_type protein
##residues 461-475 ##label MAT

REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.;
Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell
epitopes.

#cross-references MUID:89093918
#contents annotation; epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
Laureto, P.P.; DasGupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter
release by proteolytic cleavage of synaptobrevin.

#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.;
Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of
tetanus neurotoxin.

#cross-references MUID:95262688
#accession S69348
##molecule_type protein
##residues 2-31 ##label DEF

COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form
light (fragment A) and heavy (fragment B.C) chains that are
covalently linked by an interchain disulfide bond (the individual
chains are not toxic when separated). The amino end of the heavy
chain (fragment B) can be separated from the carboxyl end
(fragment C) by papain.

COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds
to gangliosides and may target the toxin to the motor end plate.
COMMENT Fragment A is a zinc-dependent endopeptidase.

COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is
internalized, and moves by retrograde transport up the axon into
the spinal cord where it can move between postsynaptic and
presynaptic neurons. It inhibits neurotransmitter release by
proteolytic cleavage of synaptobrevin (vesicle-associated
membrane protein 2).

FUNCTION
#description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide
bond in synaptobrevin 2

CLASSIFICATION
#superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane
protein; zinc

FEATURE
2-457
461-1315
461-864
865-1315
233,237
234
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853

Query Match 100.0%; Score 158; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 4.98e-18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 947 FNNFTVSFWLRVPKVSASHLE 967
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 2
ENTRY S33411 #type complete
TITLE botulinum neurotoxin type F - Clostridium barati
ORGANISM #formal_name Clostridium barati
DATE 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change
08-Sep-1997

ACCESSIONS S33411; S31860
REFERENCE S33411

#authors Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.;
Collins, M.D.; Richardson, P.T.
#journal FEMS Microbiol. Lett. (1993) 108:175-182
#title Nucleotide sequence of the gene coding for Clostridium barati
type F neurotoxin: comparison with other clostridial
neurotoxins.

#cross-references MUID:93252228
#accession S33411
##status preliminary
##molecule_type DNA
##residues 1-1268 ##label THO
##cross-references EMBL:X68262; NID:949138; PID:949139

CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
SUMMARY #length 1268 #molecular-weight 145512 #checksum 8008

Query Match 60.1%; Score 95; DB 2; Length 1268;
Best Local Similarity 64.3%; Pred. No. 3.74e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 922 YQNFVSFWVRIPK 935
QY 1 FNNFTVSFWLRVPK 14

RESULT 3
ENTRY S39791 #type complete
TITLE neurotoxin - Clostridium botulinum
ORGANISM #formal_name Clostridium botulinum
DATE 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change
24-Sep-1998

ACCESSIONS S39791
REFERENCE S39791

#authors Campbell, K.; Collins, M.D.; East, A.K.
#journal Biochim. Biophys. Acta (1993) 1216:487-491
#title Nucleotide sequence of the gene coding for Clostridium
botulinum (Clostridium argentinense) type G neurotoxin:
genealogical comparison with other clostridial neurotoxins.

#cross-references MUID:94092745
#accession S39791
##status preliminary
##molecule_type DNA
##residues 1-1297 ##label CAM
##cross-references EMBL:X74162; NID:9441275; PID:9441276

CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
SUMMARY #length 1297 #molecular-weight 149147 #checksum 2891

Query Match 60.1%; Score 95; DB 2; Length 1297;
Best Local Similarity 38.1%; Pred. No. 3.74e-05;
Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

Db 930 FDNFSINFWVRTPKYNNDIQ 950
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
ENTRY S4810 #type fragment
TITLE neurotoxin type F - Clostridium botulinum (fragment)
ORGANISM #formal_name Clostridium botulinum
DATE 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change
12-Jun-1998

```
ACCESSIONS      S48110
REFERENCE        S48103
#authors        Campbell, K.D.; Collins, M.D.; East, A.K.
#journal        J. Clin. Microbiol. (1993) 31:2255-2262
#title          Gene probes for identification of the botulinum neurotoxin
                gene and specific identification of neurotoxin types B, E,
                and F.
#cross-references MUID:94013372
#accession      S48110
#status        preliminary; translation not shown
#molecule_type DNA
#residues      1-366 #label CAM
#cross-references EMBL:X70821; NID:g407792; PID:g407793
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS       neurotoxin
SUMMARY        #length 366 #checksum 556

Query Match      58.9%; Score 93; DB 2; Length 366;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches          8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 297 YQNFISFWVRPK 310
QY      1 FNNFTVSFWLRVPK 14

RESULT 5
ENTRY   S48109 #type fragment
TITLE   neurotoxin type F - Clostridium botulinum (fragment)
ORGANISM #formal_name Clostridium botulinum
DATE     12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change
ACCESSIONS
REFERENCE S48109
#authors  Campbell, K.D.; Collins, M.D.; East, A.K.
#journal  J. Clin. Microbiol. (1993) 31:2255-2262
#title    Gene probes for identification of the botulin neurotoxin
                gene and specific identification of neurotoxin types B, E,
                and F.
#cross-references MUID:94013372
#accession      S48109
#status        preliminary; nucleic acid sequence not shown;
                translation not shown
#molecule_type DNA
#residues      1-369 #label CAM
#cross-references EMBL:X70820; NID:g407790; PID:g407791
#note          the nucleotide sequence was submitted to the EMBL Data
                Library, January 1993
CLASSIFICATION #superfamily tetanus toxin
SUMMARY        #length 369 #checksum 6830

Query Match      58.9%; Score 93; DB 2; Length 369;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches          8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 297 YQNFISFWVRPK 310
QY      1 FNNFTVSFWLRVPK 14

RESULT 6
ENTRY   I40813 #type complete
TITLE   neurotoxin type F - Clostridium botulinum
ORGANISM #formal_name Clostridium botulinum
DATE     15-Aug-1996 #sequence_revision 16-Aug-1996 #text_change
ACCESSIONS
REFERENCE I40813; S48108
#authors  East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.;
                Roberts, T.A.; Thompson, D.E.
#journal  FEMS Microbiol. Lett. (1992) 96:225-230
#title    Sequence of the gene encoding type F neurotoxin of
                Clostridium botulinum.

#accession      I40813
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      1-1274 #label RES
#cross-references GB:M92906; NID:g144866; PID:g144867
REFERENCE      S48103
#authors      Campbell, K.D.; Collins, M.D.; East, A.K.
#journal      J. Clin. Microbiol. (1993) 31:2255-2262
#title        Gene probes for identification of the botulin neurotoxin
                gene and specific identification of neurotoxin types B, E,
                and F.
#cross-references MUID:94013372
#accession      S48108
#status        preliminary; translation not shown
#molecule_type DNA
#residues      634-1002 #label CAM
#cross-references EMBL:X70816; NID:g407788; PID:g407789
KEYWORDS       neurotoxin
SUMMARY        #length 1274 #molecular-weight 146708 #checksum 2696

Query Match      58.9%; Score 93; DB 2; Length 1274;
Best Local Similarity 57.1%; Pred. No. 8.82e-05;
Matches          8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 930 YQNFISFWVRPK 943
QY      1 FNNFTVSFWLRVPK 14

RESULT 7
ENTRY   I40631 #type complete
TITLE   non-proteolytic botulinum neurotoxin type B precursor -
                Clostridium botulinum
ORGANISM #formal_name Clostridium botulinum
DATE     12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
ACCESSIONS
REFERENCE I40631; S48103; S48104; S36015
#authors  Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
#journal  Curr. Microbiol. (1994) 28:101-110
#title    Nucleotide sequence of the gene coding for non-proteolytic
                Clostridium botulinum type B neurotoxin: comparison with
                other clostridial neurotoxins.
#cross-references MUID:94122659
#accession      I40631
#status        preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues      1-1291 #label RES
#cross-references EMBL:X71343; NID:g296148; PID:g296149
REFERENCE      S48103
#authors      Campbell, K.D.; Collins, M.D.; East, A.K.
#journal      J. Clin. Microbiol. (1993) 31:2255-2262
#title        Gene probes for identification of the botulin neurotoxin
                gene and specific identification of neurotoxin types B, E,
                and F.
#cross-references MUID:94013372
#accession      S48103
#status        preliminary; nucleic acid sequence not shown;
                translation not shown
#molecule_type DNA
#residues      634-761, 'E', 763-841, 'M', 843, 'I', 845, 'N', 847-994 #label
                CAM1
#cross-references EMBL:X70814; NID:g407778; PID:g407779
#experimental_source non-proteolytic strain 2129B (Scott)
#note          the nucleotide sequence was submitted to the EMBL Data
                Library, January 1993
#accession      S48104
#status        preliminary
#molecule_type DNA
#residues      634-843, 'I', 845, 'N', 847-994 #label CAM2
#cross-references EMBL:X70819; NID:g407780; PID:g407781
#experimental_source non-proteolytic strain Eklund 2B (Colworth 229)
```


COMMENT Botulinum neurotoxin type B in these strains may posses a capable catalytic site but are nonetheless inactive.

GENETICS #gene bont/b

CLASSIFICATION #superfamily tetanus toxin

KEYWORDS metalloprotein; neurotoxin; transmembrane protein; zinc

FEATURE 2-441

442-1291 #product botulinum neurotoxin type B light chain #status predicted #label LGHT\

230,234 #product botulinum neurotoxin type B heavy chain #status predicted #label HVY\

SUMMARY #binding_site zinc (His) #status predicted\

231 #active_site Glu #status predicted

#length 1291 #molecular-weight 150513 #checksum 5240

Query Match 58.9%; Score 93; DB 2; Length 1291;

Best Local Similarity 64.3%; Pred. No. 8.82e-05;

Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 923 FLDPSVSWTRIPK 936

Cy 1 FNNFTVSFWLRVPK 14

RESULT 8

ENTRY A48940 #type complete

TITLE bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum

ALTERNATE_NAMES botulinum neurotoxin type B (Bont/B)

ORGANISM #formal_name Clostridium botulinum

DATE 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 15-May-1998

ACCESSIONS A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574

REFERENCE A48940

#authors Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.

#journal Appl. Environ. Microbiol. (1992) 58:2345-2354

#title Molecular cloning of the Clostridium botulinum structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence.

#cross-references MUID:92384550

#accession A48940

#status preliminary

##molecule_type DNA

##residues 1-1291 ##label WHE

##cross-references GB:M61186; MID:g144734; PID:g144735

##experimental_source type B, Danish

##note sequence extracted from NCBI backbone (NCBIN:112080, NCBI:P:112081); this publication is not cited in GenBank entry CLOBOTB, release 103

REFERENCE S48103

#authors Campbell, K.D.; Collins, M.D.; East, A.K.

#journal J. Clin. Microbiol. (1993) 31:2255-2262

#title Gene probes for identification of the botulinal neurotoxin gene and specific identification of neurotoxin types B, E, and F.

#cross-references MUID:94013372

#accession S48105

#status preliminary

##molecule_type DNA

##residues 634-994 ##label CAM

##cross-references EMBL:X70817; MID:g407782; PID:g407783

##experimental_source proteolytic type B, strain NCTC 7273

REFERENCE S21575

#authors Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.

#submission Submitted to the EMBL Data Library/, April 1992

#description Partial amino acid sequence of botulinum neurotoxin type B and comparison to other Clostridial neurotoxins.

#accession S21575

##molecule_type DNA

##residues 36-217,'G',219-224,'S',226-246 ##label SZA

##cross-references EMBL:Z11934; MID:g40383; PID:g40384

REFERENCE A42871

#authors Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K.; Poulain, B.; Tauc, L.; Niemann, H.

#journal J. Biol. Chem. (1992) 267:14721-14729

#title Minimal essential domains specifying toxicity of the light chains of tetanus toxin and botulinum neurotoxin type A.

#cross-references MUID:92340509

#accession A42871

#status nucleic acid sequence not shown

##molecule_type mRNA

##residues 1-313,'S',315-451 ##label KUR

##experimental_source strain Okra

##note sequence extracted from NCBI backbone (NCBI:P:109365)

REFERENCE S07155

#authors Dasgupta, B.R.; Datta, A.

#journal Biochimie (1988) 70:811-817

#title Botulinum neurotoxin type B (strain 657): partial sequence and similarity with tetanus toxin.

#cross-references MUID:89000987

#accession S07155

##molecule_type protein

##residues 2-29,'M',31-45 ##label DAS

#accession S08562

##molecule_type protein

##residues 442-463,'R',465-467 ##label DA2

REFERENCE S07128

#authors Schmidt, J.J.; Sathyamoorthy, V.; Dasgupta, B.R.

#journal Arch. Biochem. Biophys. (1985) 238:544-548

#title Partial amino acid sequences of botulinum neurotoxins types B and E.

#cross-references MUID:85197963

#accession S07128

#status preliminary

##molecule_type protein

##residues 2-16 ##label SCH1

#accession S08573

#status preliminary

##molecule_type protein

##residues 2-17 ##label SCH2

#accession S08574

#status preliminary

##molecule_type protein

##residues 442-459 ##label SCH3

REFERENCE S27125

#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; Dasgupta, B.R.; Montecucco, C.

#journal Nature (1992) 359:832-835

#title Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.

#cross-references MUID:93063293

#contents annotation

COMMENT Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses. This toxin is activated by cleavage into two chains linked by a disulfide bond.

GENETICS #gene bont/b

FUNCTION #description catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2

CLASSIFICATION #superfamily tetanus toxin

KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc

FEATURE 2-441

442-1291 #product bontoxilysin B light chain #status experimental

230,234 #product bontoxilysin B heavy chain #status experimental

SUMMARY #binding_site zinc (His) #status predicted\

231 #active_site Glu #status predicted

#length 1291 #molecular-weight 150801 #checksum 9744

Query Match 58.9%; Score 93; DB 1; Length 1291;

Best Local Similarity 64.3%; Pred. No. 8.82e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 923 FLDFSVSFWIRPK 936
|:|:|:|:|:|:|:
QY 1 FNNFTVSEFWLRVPK 14

RESULT 9
ENTRY BTCLAB #type complete
TITLE bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum

ALTERNATE_NAMES botulinum neurotoxin type A
ORGANISM #formal_name Clostridium botulinum
DATE 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 15-May-1998
A35294; S09492; S68220; A33401; A53884; A60025; A27000
A35294
Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. (1990) 265:9153-9158
#journal The complete sequence of botulinum neurotoxin type A and comparison with other clostridial neurotoxins.
#title #cross-references MUID:90264400
#accession A35294
##molecule_type DNA
##residues 1-1296 #label BIN
##cross-references GB:M30196; NID:g144864; PID:g144865
##experimental_source strain 62A, subtype A
S09492

REFERENCE
#authors Thompson, D.E.; Brehm, J.K.; Oultam, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.; Melling, J.; Minton, N.P.
#journal Eur. J. Biochem. (1990) 189:73-81
#title The complete amino acid sequence of the Clostridium botulinum type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene.
#cross-references MUID:90235864
#accession S09492
##molecule_type DNA
##residues 1, 'Q', 3-26, 'V', 28-1296 #label THO
##cross-references EMBL:X52066; NID:g40381; PID:g40382
##experimental_source NCTC 2916

REFERENCE
#authors Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
#journal FEBS Lett. (1995) 376:41-44
#title Molecular characterization of two forms of nontoxic-nonhemagglutinin components of Clostridium botulinum type A progenitor toxins.
#cross-references MUID:96096783
#accession S68220
##status preliminary
##molecule_type DNA
##residues 1-12 #label FUJ
##cross-references EMBL:D67030; DDBJ:D50421; NID:g2160224
A33401

REFERENCE
#authors Betley, M.J.; Somers, E.; DasGupta, B.R.
#journal Biochem. Biophys. Res. Commun. (1989) 162:1388-1395
#title Characterization of botulinum type A neurotoxin gene: delineation of the N-terminal encoding region.
#cross-references MUID:89350959
#accession A33401
##molecule_type DNA
##residues 1-35 #label BEI
##cross-references GB:M27892; NID:g144880; PID:g551776
A53884

REFERENCE
#authors Gimenez, J.A.; DasGupta, B.R.
#journal J. Protein Chem. (1993) 12:351-363
#title Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, and 18 kD fragments.
#cross-references MUID:94000342
#accession A53884
#status preliminary

##molecule_type protein
##residues 867-880;1148-1217,'Y',1219 #label GIM
##experimental_source strain Hall
##note sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after extraction from NCBI backbone

REFERENCE
#authors DasGupta, B.R.; Dekleva, M.L.
#journal Biochimie (1990) 72:661-664
#title Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and around the nicking site.
#cross-references MUID:91120847
#accession A60025
##molecule_type protein
##residues 2-6;445-453,'X',455-457 #label DAS1
A27000
DasGupta, B.R.; Foley, J.; Niece, R.
#journal Biochemistry (1987) 26:4162
#title Partial sequence of the light chain of botulinum neurotoxin type A.
#accession A27000
##molecule_type protein
##residues 2-47 #label DAS2
A49708
Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.; Niemann, H.
#journal J. Biol. Chem. (1994) 269:1617-1620
#title Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.
#cross-references MUID:94124495
#contents annotation
#comment Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synapses. This toxin is activated by cleavage into two chains linked by a disulfide bond.

GENETICS
#gene atx; bota
FUNCTION
#description catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated 25K protein (SNAP-25)
KEYWORDS
#superfamily tetanus toxin
#disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
FEATURE
2-444
#product bontoxilysin A light chain #status experimental
#label LGHT\

445-1296
#product bontoxilysin A heavy chain #status experimental
#label HVY\
223,227
#binding_site zinc (His) #status predicted\
224
#active_site Glu #status predicted
SUMMARY
#length 1296 #molecular-weight 149425 #checksum 7102

Query Match 57.0%; Score 90; DB 1; Length 1296;
Best Local Similarity 57.1%; Pred. No. 3.15e-04;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 938 YENFSTSEFWIRPK 951
::|:|:|:|:|:|:|:
QY 1 FNNFTVSEFWLRVPK 14

RESULT 10
ENTRY S48106 #type fragment
TITLE neurotoxin type E - Clostridium botulinum (fragment)
ORGANISM #formal_name Clostridium botulinum
DATE 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 12-Jun-1998
S48106
S48103
REFERENCE
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulinum neurotoxin gene and specific identification of neurotoxin types B, E, and F.
#cross-references MUID:94013372

```

#accession      S48106
#status          preliminary; nucleic acid sequence not shown;
                  translation not shown
#molecule_type DNA
##residues      1-367 ##label CAM
##cross-references EMBL:X70818; NID:g407784; PID:g407785
                  the nucleotide sequence was submitted to the EMBL Data
                  Library, January 1993
CLASSIFICATION  #superfamily tetanus toxin
KEYWORDS        neurotoxin
SUMMARY         #length 367 #checksum 184

Query Match
Best Local Similarity 55.7%; Score 88; DB 2; Length 367;
Matches 7; Conservativity 7; Mismatches 0; Indels 0; Gaps 0;

Db 297 YKNFSISFWVRIPN 310
QY 1 FNNFTVSFWLRVPK 14
::|::|::|::|::|:

RESULT 11
ENTRY JH0257 #type complete
TITLE botulinum neurotoxin type E precursor - Clostridium botulinum
        (strain Beluga)
ORGANISM #formal_name Clostridium botulinum
DATE 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
        08-Sep-1997
ACCESSION JH0257; B35294; A60027
#authors Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
#journal Biochem. Biophys. Res. Commun. (1992) 183:107-113
#title Sequences of the botulinum neurotoxin E derived from
        Clostridium botulinum type E (strain Beluga) and
        Clostridium butyricum (strains ATCC 43181 and ATCC 43755).
#cross-references MUID:92181428
#accession JH0257
##status nucleic acid sequence not shown
##molecule_type DNA
##residues 1-1251 ##label POU
##cross-references EMBL:X62089; NID:g40393; FID:g40394
REFERENCE A35294
#authors Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Vernars, K.;
        Niemann, H.
#journal J. Biol. Chem. (1990) 265:9153-9158
#title The complete sequence of botulinum neurotoxin type A and
        comparison with other clostridial neurotoxins.
#cross-references MUID:90264400
#accession B35294
##status not compared with conceptual translation
##molecule_type DNA
##residues 1-197, 'S', 199-252 ##label BIR
REFERENCE A60027
#authors Gimenez, J.A.; DasGupta, B.R.
#journal Biochimie (1990) 72:213-217
#title Botulinum neurotoxin type E fragmented with endoproteinase
        Lys-C reveals the site trypsin nicks and homology with
        tetanus neurotoxin.
#cross-references MUID:90344918
#accession A60027
##molecule_type protein
##residues 420-427 ##label GIM
##note this fragment was generated by proteolysis with Lys-C
        rather than with trypsin
COMMENT The clostridial neurotoxins are highly potent protein toxins that
        inhibit neurotransmitter release at various synapses.
        The heavy chain mediates the binding of toxin to cell receptors
        while the light chain appears to enter target cells.
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS        neurotoxin
FEATURE
2-427 #product botulinum neurotoxin light chain (toxin
        fragment A) #status predicted #label LIG\

```

```

423-1251      #product botulinum neurotoxin heavy chain (toxin
412-426      fragment BC) #status predicted #label HEA\
SUMMARY      #disulfide_bonds #status predicted
              #length 1251 #molecular-weight 143843 #checksum 3754

Query Match      55.7%; Score 88; DB 2; Length 1251;
Best Local Similarity 50.0%; Pred. No. 7.30e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db      912 YKNFSISFWVRIPN 925
QY      1 FNNFTVSFWLRVPK 14
          ::||::|||:|:|:

RESULT      12
ENTRY      JH0256      #type complete
TITLE      botulinum neurotoxin type E precursor - Clostridium butyricum
ORGANISM   #formal_name Clostridium butyricum
DATE       30-Jun-1992 #sequence_revision 15-May-1998 #text_change
              15-May-1998
ACCESSIONS JH0256; S16145
REFERENCE   JH0256
            Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
            Biochem. Biophys. Res. Commun. (1992) 183:107-113
            Sequences of the botulinum neurotoxin E derived from
            Clostridium botulinum type E (strain Beluga) and
            Clostridium butyricum (strains ATCC 43181 and ATCC 43755).
            #cross-references MUID:92181428
            #accession JH0256
            ##status nucleic acid sequence not shown
            ##molecule_type DNA
            ##residues 1-27, 'E', 29-1251 ##label POU
            ##cross-references EMBL:X62088; NID:g40379
            ##experimental_source strains ATCC 43181 and ATCC 43755
REFERENCE   S16145
            Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.;
            Tsuzuki, K.; Yokosawa, N.; Oguma, K.
            J. Gen. Microbiol. (1991) 137:519-525
            Cloning of a DNA fragment encoding the 5'-terminus of the
            botulinum type E toxin gene from Clostridium butyricum
            strain BL6340.
            #cross-references MUID:91237316
            #accession S16145
            ##status preliminary
            ##molecule_type DNA
            ##residues 1-229, 'M', 231-252 ##label FUJ
            ##cross-references EMBL:X53180; NID:g40407; PID:g40408
            ##experimental_source strain BL6340
COMMENT     The clostridial neurotoxins are toxins that inhibit
            neurotransmitter release at synaptic junctions.
COMMENT     The heavy chain mediates the binding of toxin to cell receptors
            while the light chain appears to enter target cells.
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS      neurotoxin
FEATURE       2-422      #product botulinum neurotoxin type E light chain #status
                        predicted #label LIG\
                        423-1251      #product botulinum neurotoxin type E heavy chain #status
                        predicted #label HEA\
                        412-426      #disulfide_bonds #status predicted
SUMMARY      #length 1251 #molecular-weight 143396 #checksum 3055

Query Match      55.7%; Score 88; DB 2; Length 1251;
Best Local Similarity 50.0%; Pred. No. 7.30e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db      912 YKNFSISFWVRIPN 925
QY      1 FNNFTVSFWLRVPK 14
          ::||::|||:|:|:

RESULT      13
ENTRY      S21178      #type complete

```


TITLE botulinum neurotoxin type E precursor - Clostridium botulinum
ORGANISM #formal_name Clostridium botulinum
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
ACCESSIONS 12-Jun-1998
REFERENCE S21178; S48107; S18111
#authors Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.;
Minton, N.P.
#journal Eur. J. Biochem. (1992) 204:657-667
#title The complete amino acid sequence of the Clostridium botulinum
type-E neurotoxin, derived by nucleotide-sequence analysis
of the encoding gene.
#cross-references MUID:92174922
#accession S21178
#molecule_type DNA
#residues 1-1252 #label WHE
#cross-references EMBL:X62683; NID:g40397; PID:g40398
REFERENCE S48103
#authors Campbell, K.D.; Collins, M.D.; East, A.K.
#journal J. Clin. Microbiol. (1993) 31:2255-2262
#title Gene probes for identification of the botulinum neurotoxin
gene and specific identification of neurotoxin types B, E,
and F.
#cross-references MUID:94013372
#accession S48107
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 616-982 #label CAM
#cross-references EMBL:X70815; NID:g407786; PID:g407787
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
FEATURE
1-422 #product botulinum neurotoxin type E light chain #status
423-1252 predicted #label LCH\
#product botulinum neurotoxin type E heavy chain #status
412-426 predicted #label HCH\
#disulfide_bonds #status predicted
SUMMARY #length 1252 #molecular-weight 143637 #checksum 7136
Query Match 55.7%; Score 88; DB 2; Length 1252;
Best Local Similarity 50.0%; Pred. No. 7.30e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;
Db 912 YKNFSISFWIRPN 925
QY 1 FNNFTVSFWLRVPK 14
RESULT 14
ENTRY I40645 #type complete
TITLE botulinum neurotoxin type A - Clostridium botulinum
ORGANISM #formal_name Clostridium botulinum
DATE 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change
16-Feb-1997
ACCESSIONS I40645
REFERENCE I40645
#authors Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
#journal Res. Microbiol. (1993) 144:547-556
#title Sequence of the gene coding for the neurotoxin of Clostridium
botulinum type A associated with infant botulism:
comparison with other clostridial neurotoxins.
#cross-references MUID:94143603
#accession I40645
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-1296 #label RES
#cross-references EMBL:X73423; NID:g507070; PID:g507071
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS neurotoxin
SUMMARY #length 1296 #molecular-weight 149410 #checksum 1997

Query Match 53.8%; Score 85; DB 2; Length 1296;
Best Local Similarity 50.0%; Pred. No. 2.53e-03;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;
Db 938 YENFSTFWIKRPK 951
QY 1 FNNFTVSFWLRVPK 14
RESULT 15
ENTRY MNVWC #type complete
TITLE nonstructural protein NS - tomato spotted wilt virus (strain
CPNH1)
ORGANISM #formal_name tomato spotted wilt virus
DATE 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change
05-Sep-1997
ACCESSIONS JQ0547
REFERENCE JQ0547
#authors de Haan, P.; Wagemaekers, L.; Peters, D.; Goldbach, R.
#journal J. Gen. Virol. (1990) 71:1001-1007
#title The S RNA segment of tomato spotted wilt virus has an
ambisense character.
#cross-references MUID:90264829
#accession JQ0547
#molecule_type genomic RNA
#residues 1-464 #label DEH
#cross-references DDBJ:D00645; NID:g222685; PID:d1000395; PID:g222686
GENETICS NS
#gene NS
#map_position segment S
CLASSIFICATION #superfamily tomato spotted wilt virus nonstructural protein
NS
KEYWORDS glycoprotein; nonstructural protein
FEATURE
132,210,270,291, #binding_site carbohydrate (Asn) (covalent) #status
381 predicted
SUMMARY #length 464 #molecular-weight 52448 #checksum 2559
Query Match 53.2%; Score 84; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 3.82e-03;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 239 SHEKLSLWLRVPK 252
QY 2 NNFTVSFWLRVPK 15

Search completed: Wed Aug 4 15:25:22 1999
Job time : 7 secs.

This Page Blank (uspto)

M P R C H

(TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPrch_pp Protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:25:40 1999; MasPar time 3.28 Seconds
Tabular output not generated. 181.151 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSFWLRPKVVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 30.160; Variance 44.662; scale 0.675

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	158	100.0	1314	1	TETX_CLOTE	TETANUS TOXIN PRECURSOR	2.29e-20
2	95	60.1	1296	1	BXG_CLOBO	BOTULINUM NEUROTOXIN T	4.20e-05
3	93	58.9	1274	1	BXF_CLOBO	BOTULINUM NEUROTOXIN T	1.08e-05
4	93	58.9	1290	1	BXB_CLOBO	BOTULINUM NEUROTOXIN T	1.08e-05
5	90	57.0	1295	1	BXA_CLOBO	BOTULINUM NEUROTOXIN T	4.41e-05
6	88	55.7	1250	1	BXE_CLOBO	BOTULINUM NEUROTOXIN T	1.11e-04
7	88	55.7	1250	1	BXE_CLOBO	BOTULINUM NEUROTOXIN T	1.11e-04
8	87	55.7	1051	1	VP2_AHSV6	OUTER CAPSID PROTEIN V	1.76e-04
9	84	53.2	464	1	VNSX_TSWV1	NONSTRUCTURAL PROTEIN	6.88e-04
10	84	53.2	467	1	VNSX_TSWV1	NONSTRUCTURAL PROTEIN	6.88e-04
11	81	51.3	1290	1	BXC1_CLOBO	BOTULINUM NEUROTOXIN T	2.63e-03
12	77	48.7	449	1	VNSX_INSVN	NONSTRUCTURAL PROTEIN	1.51e-02
13	76	48.1	1196	1	BXCN_CLOBO	BOTULINUM NEUROTOXIN T	2.32e-02
14	72	45.6	1276	1	BXD_CLOBO	BOTULINUM NEUROTOXIN T	1.25e-01
15	70	44.3	1162	1	BXEN_CLOBO	BOTULINUM NEUROTOXIN T	2.84e-01
16	70	44.3	1162	1	BXEN_CLOBO	BOTULINUM NEUROTOXIN T	2.84e-01
17	67	42.4	789	1	A12M_YEAST	PUTATIVE COX1/OX13 INT	9.49e-01
18	66	41.8	241	1	CRTA_RHOCA	SPHEROIDENE MONOOXYGEN	1.41e+00
19	63	39.9	547	1	FTSO_STRGR	CELL DIVISION PROTEIN	4.48e+00
20	63	39.9	547	1	CP78_MALZE	CYTCHROME P450 78A1 (4.48e+00
21	63	39.9	605	1	ALBU_PIG	SERUM ALBUMIN PRECURSOR	4.48e+00
22	63	39.9	728	1	YJ89_YEAST	HYPOTHETICAL 85.0 KD P	4.48e+00
23	62	39.2	581	1	AMY1_SCHPO	PROBABLE ALPHA-AMYLASE	6.53e+00

24	62	39.2	867	1	RRPO_BYDV1	PUTATIVE RNA-DIRECTED	6.53e+00
25	62	39.2	1683	1	POLG_DEN2T	GENOME POLYPROTEIN [CO	6.53e+00
26	62	39.2	3388	1	POLG_DEN2P	GENOME POLYPROTEIN [CO	6.53e+00
27	62	39.2	3390	1	POLG_DEN3	GENOME POLYPROTEIN [CO	6.53e+00
28	61	38.6	602	1	VE1_HPV15	REPLICATION PROTEIN E1	9.49e+00
29	61	38.6	711	1	YRFE_ECOLI	HYPOTHETICAL 79.5 KD P	9.49e+00
30	60	38.0	281	1	PH4H_CHRVO	PHENYLALANINE-4-HYDROX	1.37e+01
31	60	38.0	352	1	CITC_ECOLI	[CITRATE (PRO-3S)-LYAS	1.37e+01
32	60	38.0	523	1	CP78_SOYBN	CYTCHROME P450 78A3 (1.37e+01
33	60	38.0	609	1	VE1_HPV37	REPLICATION PROTEIN E1	1.37e+01
34	60	38.0	647	1	SOHC_RHISN	PROBABLE SQUALENE-HOP	1.37e+01
35	60	38.0	1437	1	DPO3_BACSU	DNA POLYMERASE III, AL	1.37e+01
36	60	38.0	1608	1	RRPL_BDV	RNA POLYMERASE BETA SU	1.37e+01
37	59	37.3	102	1	VAPA_BACNO	VIRULENCE-ASSOCIATED P	1.97e+01
38	59	37.3	237	1	LECA_DOLLA	LECTIN.	1.97e+01
39	59	37.3	371	1	RODA_HAEIN	ROD SHAPE-DETERMINING	1.97e+01
40	59	37.3	382	1	TGT_HAEIN	QUEUINE TRNA-RIBOSYLTR	1.97e+01
41	59	37.3	532	1	TYRO_RANNI	TYROSINASE PRECURSOR (1.97e+01
42	59	37.3	633	1	AGP1_YEAST	ASPARAGINE/GLUTAMINE P	1.97e+01
43	59	37.3	641	1	YGG0_YEAST	HYPOTHETICAL 73.1 KD P	1.97e+01
44	58	36.7	342	1	FPP2_LUPAL	FARNESYL PYROPHOSPHATE	2.82e+01
45	58	36.7	537	1	AREH_SCHPO	PROBABLE STEROL O-ACYL	2.82e+01

ALIGNMENTS

RESULT 1
ID TETX_CLOTE STANDARD; PRT; 1314 AA.
AC P04958;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).
OS CLOSTRIDIUM TETANI.
OG PLASMID.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87053814.
RA EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,
RA WELTER U., HUDEL M., HABERMANN E., NIEMANN H.;
RT "Tetanus toxin: primary structure, expression in E. coli, and
RT homology with botulinum toxins.";
RL EMBO J. 5:2495-2502(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CN3911;
RX MEDLINE; 87040747.
RA FAIRWEATHER N.F., LYNESS V.A.;
RT "The complete nucleotide sequence of tetanus toxin.";
RT NUCLEIC ACIDS RES. 14:7809-7812(1986).
RN [3]
RP SEQUENCE OF 742-1314 FROM N.A.
RX MEDLINE; 86085672.
RA FAIRWEATHER N.F., LYNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;
RT "Cloning, nucleotide sequencing, and expression of tetanus toxin
RT fragment C in Escherichia coli.";
RL J. BACTERIOL. 165:21-27(1986).
RN [4]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE; 90201034.
RA KRIEGLSTEIN K., HENSCHEN A., WELTER U., HABERMANN E.;
RT "Arrangement of disulfide bridges and positions of sulphydryl groups
RT in tetanus toxin.";
RL EUR. J. BIOCHEM. 188:39-45(1990).
RN [5]
RP PARTIAL SEQUENCE.
RX MEDLINE; 92037649.
RA KRIEGLSTEIN K.G., HENSCHEN A.H., WELTER U., HABERMANN E.;
RT "Limited proteolysis of tetanus toxin. Relation to activity and
RT identification of cleavage sites.";
RL EUR. J. BIOCHEM. 202:41-51(1991).

RN [6]
R2 IDENTIFICATION AS ZINC-PROTEASE.
RX MEDLINE; 93010948.
RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
RA MONTECUCCO C.;
RT "Tetanus toxin is a zinc protein and its inhibition of
RT neurotransmitter release and protease activity depend on zinc";
RL EMBO J. 11:3577-3583(1992).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 93063293.
RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
RA DASGUPTA B.R., MONTECUCCO C.;
RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT by proteolytic cleavage of synaptobrevin";
RL NATURE 359:832-835(1992).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
RX MEDLINE; 97475217.
RA UMLAND T.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
RA SAX M.;
RT "Structure of the receptor binding fragment HC of tetanus
RT neurotoxin";
RL NAT. STRUCT. BIOL. 4:788-792(1997).
CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-1-PHE-77
CC BOND OF SYNAPTOSOMAL-2.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-1-PHE-77 BOND IN
CC SYNAPTOSOMAL-2.
CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
CC AND ARE NON-TOXIC AFTER SEPARATION.
CC -1- THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO GANGLIOSIDE RECEPTORS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04436; G40770; -.
DR EMBL; M12739; G144921; -.
DR EMBL; X06214; G40774; -.
DR PIR; A25689; BTCLTN.
DR PDB; 1AF9; 29-APR-98.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC; PLASMID;
KW 3D-STRUCTURE.
FT INIT_MET 0
FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
FT METAL 232 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 233 233 BY SIMILARITY.
FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
FT TRANSMEM 226 246 POTENTIAL.
FT TRANSMEM 689 689 POTENTIAL.
FT DISULFID 438 466 INTERCHAIN.
FT DISULFID 1076 1092
SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;

Query Match 100.0%; Score 158; DB 1; Length 1314;
Best Local Similarity 100.0%; Pred. No. 2.29e-20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 946 FNNFTVSEWLRVPKVSASHLE 966
QY 1 FNNFTVSEWLRVPKVSASHLE 21
RESULT 2
ID BXG_CLOBO STANDARD; PRT; 1296 AA.
AC Q60393;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE G PRECURSOR (EC 3.4.24.69) (BONT/G)
DE (BONTOLYSIN G).
GN BOTG.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-113 / 30;
RX MEDLINE; 94092745.
RA CAMPBELL K., COLLINS M.D., EAST A.K.;
RT "Nucleotide sequence of the gene coding for Clostridium botulinum
RT (Clostridium argentineense) type G neurotoxin: genealogical comparison
RT with other clostridial neurotoxins";
RL BIOCHIM. BIOPHYS. ACTA 1216:487-491(1993).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDATE
CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X74162; G441276; -.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PFAM; PF00099; zn-protease; 1.
DR HSSP; P04958; 1AF9.
KW NEUROTOXIN; HYDROLASE; METALLOPROTEASE; ZINC.
KW INIT_MET 0
FT CHAIN 1 441 BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT CHAIN 442 1296 BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 230 233 BY SIMILARITY.
FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 435 449 INTERCHAIN (PROBABLE).
SQ SEQUENCE 1296 AA; 149013 MW; 8570B903 CRC32;

Query Match 60.1%; Score 95; DB 1; Length 1296;
Best Local Similarity 38.1%; Pred. No. 4.20e-06;
Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

```
RESULT 3
ID BXF_CLOBO STANDARD; PRT: 1274 AA.
AC P30996;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE F PRECURSOR (EC 3.4.24.69) (BONT/F)
GN (BONTOLYLISIN F).
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
RN CLOSTRIDIUM.
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 23387;
RX MEDLINE; 93012902.
RA EAST A.K., RICHARDSON P.T., ALLAWAY D., COLLINS M.D.,
RA ROBERTS T.A., THOMPSON D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum."
RL FEMS MICROBIOL. LETT. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RX STRAIN=HOBBS FT10;
RX MEDLINE; 94297488.
RA EAST A.K., COLLINS M.D.;
RI "Conserved structure of genes encoding components of botulinum
RI neurotoxin complex M and the sequence of the gene coding for the
RI nontoxic component in nonproteolytic Clostridium botulinum type F."
RL CURR. MICROBIOL. 29:69-77(1994).
RN [3]
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RI "Gene probes for identification of the botulin neurotoxin gene and
RI specific identification of neurotoxin types B, E, and F."
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 94230352.
RA YAMASAKI S., BAUMEISTER A., BINZ T., BLASI J., LINK E., CORNILLE F.,
RA ROQUES B., FYKSE E.M., STEPHOF T.C., JAHN R., NIEMANN H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulin neurotoxins and tetanus toxin."
RL J. BIOL. CHEM. 269:12764-12772(1994).
RN [1]
RP FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
RP RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
RP AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
RP WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
RP INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
RP ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-LYS-59
RP BOND OF SYNAPTOBREVIN-1 AND -2.
RN [1]
RP SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
RN A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
RN ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
RN CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
RN [1]
RP SUBCELLULAR LOCATION: SECRETED.
RN [1]
RP THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
RN NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
RN [1]
RP SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
RN METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
RN SUBFAMILY.
RN [1]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [1]
RN EMBL; M92906; G144867; -.
```

```
DR EMBL; S73676; G765179; -
DR EMBL; X70820; G407791; -
DR EMBL; X70816; G407789; -
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PFAM; PF00099; zn-protease; 1.
DR HSSP; P04958; JAF9.
KM NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT CHAIN 1 436
FT CHAIN 437 1274
FT METAL 227 227
FT ACT SITE 228 228
FT METAL 231 231
FT DISULFID 429 445
SQ SEQUENCE 1274 AA; 146709 MW; 10954D22 CRC32;
Query Match
Best Local Similarity 58.98; Score 93; DB 1; Length 1274;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
Db 930 YQNFISFWVRIPK 943
QY 1 FNNFTVFWLRVPEK 14
RESULT 4
ID BXF_CLOBO STANDARD; PRT: 1290 AA.
AC P10844; P10843;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B PRECURSOR (EC 3.4.24.69) (BONT/B)
DE (BONTOLYLISIN B).
GN BOTB.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
RN CLOSTRIDIUM.
RP [1]
RC SEQUENCE FROM N.A.
RX MEDLINE; 92384550.
RA WHELAN S.M., ELMORE M.J., BODSWORTH N.J., BREHM J.K., ATKINSON T.,
RA MINTON N.P.;
RT "Molecular cloning of the Clostridium botulinum structural gene
RT encoding the type B neurotoxin and determination of its entire
RT nucleotide sequence."
RL APPL. ENVIRON. MICROBIOL. 58:2345-2354(1992).
RN [2]
RP SEQUENCE OF 35-245 FROM N.A.
RC STRAIN=NCTC 7273;
RA SZABO E.A., PEMBERTON J.M., DESMARCHELIER P.M.;
RL SUBMITTED (APR-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 633-993 FROM N.A.
RC STRAIN=NCTC 7273;
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RI "Gene probes for identification of the botulin neurotoxin gene and
RI specific identification of neurotoxin types B, E, and F."
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-44 AND 441-466.
RC STRAIN=657;
RX MEDLINE; 89000987.
RA DASGUPTA B.R., DATTA A.;
RI "Botulinum neurotoxin type B (strain 657): partial sequence and
RI similarity with tetanus toxin."
RL BIOCHIMIE 70:811-817(1988).
RN [5]
RP SEQUENCE OF 1-16 AND 441-458.
RC STRAIN=OKRA;
RX MEDLINE; 85197963.
RA SCHMIDT J.J., SATHYAMOORTHY V., DASGUPTA B.R.;
RI "Partial amino acid sequences of botulinum neurotoxins types B and
RI E."
RN [1]
RN EMBL; M92906; G144867; -.
```



```

RN  ARCH. BIOCHEM. BIOPHYS. 238:544-548(1985).
RP  [6]
RX  IDENTIFICATION AS ZINC-PROTEASE.
RA  MEDLINE; 93054694.
RT  SCHIAVO G., ROSSETTO O., SANTUCCI A., DASGUPTA B.R., MONTECUCCO C.;
RN  "Botulinum neurotoxins are zinc proteins.";
    J. BIOL. CHEM. 267:23479-23483(1992).
RP  [7]
RX  IDENTIFICATION OF SUBSTRATE.
RA  MEDLINE; 93063293.
RT  SCHIAVO G., BENENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
RN  DASGUPTA B.R., MONTECUCCO C.;
    "Tetanus and botulinum-B neurotoxins block neurotransmitter release
    by proteolytic cleavage of synaptobrevin.";
    NATURE 359:832-835(1992).
RL  -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC  RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC  AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC  WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC  INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC  ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF
CC  SYNAPTOSOMAL-2.
CC  -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC  A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC  ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC  CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC  -1- SUBCELLULAR LOCATION: SECRETED.
CC  -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC  NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC  METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC  SUBFAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M81186; G144735; -.
DR  EMBL; Z11934; G40384; -.
DR  EMBL; X70817; G407783; -.
DR  PIR; S07128; S07128.
DR  PIR; S07155; S07155.
DR  PIR; S08562; S08562.
DR  PIR; S08573; S08573.
DR  PIR; S08574; S08574.
DR  PIR; A48940; A48940.
DR  PROSITE; PS00142; ZINC_PROTEASE; 1.
DR  PFM; PF00099; zn-protease; 1.
DR  HSSP; P04958; 1A9.
KW  NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT  INIT_MET 0 0
FT  CHAIN 1 440
FT  CHAIN 1 1290
FT  METAL 229 229
FT  ACT_SITE 230 230
FT  METAL 233 233
FT  DISULFID 436 445
FT  CONFLICT 29 29
FT  CONFLICT 217 217
FT  CONFLICT 224 224
FT  CONFLICT 463 463
SQ  SEQUENCE 1290 AA; 150670 MW; 479BBC70 CRC32;
    BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
    BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
    ZINC (CATALYTIC) (BY SIMILARITY).
    BY SIMILARITY.
    ZINC (CATALYTIC) (BY SIMILARITY).
    INTERCHAIN (PROBABLE).
    T -> M (IN REF. 4).
    R -> G (IN REF. 2).
    A -> S (IN REF. 2).
    S -> R (IN REF. 4).
Query Match 58.9%; Score 93; DB 1; Length 1290;
Best Local Similarity 64.3%; Pred. No. 1.08e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```

QY	1	ENNETVSEFWLRVPK 14
RESULT	5	
ID	BXA.CLOBO	STANDARD; PRT; 1295 AA.
AC	P10845; P18639; P01561;	
DT	01-JUL-1989 (REL. 11, CREATED)	
DT	01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)	
DE	BOTULINUM NEUROTOXIN TYPE A PRECURSOR (EC 3.4.24.69) (BONT/A)	
DE	(BONTOLYSIN A).	
GN	BOTA OR BNA.	
OS	CLOSTRIDIUM BOTULINUM.	
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;	
OC	CLOSTRIDIUM.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=NCTC 2916;	
RX	MEDLINE; 90235864.	
RA	THOMPSON D.E., BREHM J.K., OULTRAM J.D., SWINFELD T.-C.,	
RA	SHONE C.C., ATKINSON T., MELLING J., MINTON N.P.;	
RT	"The complete amino acid sequence of the Clostridium botulinum type A	
RT	neurotoxin, deduced by nucleotide sequence analysis of the encoding	
RT	gene.";	
RL	EUR. J. BIOCHEM. 189:73-81(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=62A;	
RX	MEDLINE; 90264400.	
RA	BINZ B., KUARZONO H., WILLE M., FREVENT J., WERNARS K., NIEMANN H.;	
RT	"The complete sequence of botulinum neurotoxin type A and comparison	
RT	with other clostridial neurotoxins.";	
RL	J. BIOL. CHEM. 265:9153-9158(1990).	
RN	[3]	
RP	SEQUENCE OF 1-34 FROM N.A.	
RC	STRAIN=HALL;	
RX	MEDLINE; 89350959.	
RA	BETLEY M.J., SOMERS E., DASGUPTA B.R.;	
RT	"Characterization of botulinum type A neurotoxin gene: delineation of	
RT	the N-terminal encoding region.";	
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 162:1388-1395(1989).	
RN	[4]	
RP	SEQUENCE OF 1-16.	
RC	MEDLINE; 84178501.	
RA	SCHMIDT J.J., SARTYMOORTHY V., DASGUPTA B.R.;	
RT	"Partial amino acid sequence of the heavy and light chains of	
RT	botulinum neurotoxin type A.";	
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 119:900-904(1984).	
RN	[5]	
RP	SEQUENCE OF 1-46.	
RA	DASGUPTA B.R., FOLEY J., NIECE R.;	
RT	"Partial sequence of the light chain of botulinum neurotoxin type A.";	
RL	BIOCHEMISTRY 26:4162-4162(1987).	
RN	[6]	
RP	SEQUENCE OF 1-5 AND 444-456.	
RX	MEDLINE; 91120847.	
RA	DASGUPTA B.R., DEKLEVA M.L.;	
RT	"Botulinum neurotoxin type A: sequence of amino acids at the	
RT	N-terminus and around the nicking site.";	
RL	BIOCHIMIE 72:661-664(1990).	
RN	[7]	
RP	SEQUENCE OF 448-464 AND 872-895.	
RX	MEDLINE; 89024662.	
RA	SATHMOORTHY V., DASGUPTA B.R., FOLEY J., NIECE R.L.;	
RT	"Botulinum neurotoxin type A: cleavage of the heavy chain into two	
RT	halves and their partial sequences.";	
RL	ARCH. BIOCHEM. BIOPHYS. 266:142-151(1988).	
RN	[8]	
RP	SEQUENCE OF 448-482.	
RX	MEDLINE; 85285016.	
RA	SHONE C.C., HAMBLETON P., MELLING J.;	
RT	"Inactivation of Clostridium botulinum type A neurotoxin by trypsin	
RT	and purification of two tryptic fragments. Proteolytic action near	

RT the COOH-terminus of the heavy subunit destroys toxin-binding
RT activity.";
RL EUR. J. BIOCHEM. 151:75-82(1985).
RN [9]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 94063091.
RA SCHIAVO G., SANTUCCI A., DASGUPTA B.R., MEHTA P.P., JONTES J.,
RA BENFENATI F., WILSON M.C., MONTECUCCHI C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS LETT. 335:99-103(1993).
RN [10]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 94124495.
RA BINZ T., BLASI J., YAMASAKI S., BAUMEISTER A., LINK E., SUDHOF T.C.,
RA JAHN R., NIEMANN H.;
RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
RL J. BIOL. CHEM. 269:1617-1620(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE. IT CATALYZES THE HYDROLYSIS OF THE 197-GLN-1-ARG-
CC 198 BOND IN SNAP-25.
CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).
CC -----
DR EMBL; X52066; G40382;
DR EMBL; M30196; G144865;
DR EMBL; M27892; G551776;
DR PIR; A35294; BTCLAB.
DR PIR; S09492; S09492.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PFAM; PFC0099; zn-protease; 1.
DR HSSP; P04958; IAF9.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT INIT_MET 0
FT CHAIN 1
FT CHAIN 447
FT CHAIN 1295
FT METAL 222
FT ACT_SITE 223
FT METAL 226
FT DISULFID 429
FT TRANSMEM 626
FT TRANSMEM 646
FT VARIANT 655
FT VARIANT 26
FT CONFLICT 1
FT CONFLICT 479
FT CONFLICT 875
FT CONFLICT 891
SQ SEQUENCE 1295 AA; 149322 MW; 2E333BFB CRC32;
Query Match 57.0%; Score 90; DB 1; Length 1295;
Best Local Similarity 57.1%; Pred. No. 4.41e-05;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 937 YENSTSWIRIPK 950
QY 1 FNNFTVSEWLRVPK 14
RESULT 6
ID BXE_CLOBO STANDARD; PRT; 1250 AA.
AC Q00496;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE E PRECURSOR (EC 3.4.24.69) (BONT/E)
DE (BONTXILYSIN E).
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BEUGA;
RX MEDLINE; 92181428.
RA POULET S., HAUSER D., QUANZ M., NIEMANN H., POPOFF M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:107-113(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92174922.
RA WHELAN S.M., ELMORE M.J., BODSWORTH N.J., ATKINSON T., MINTON N.P.;
RT "The complete amino acid sequence of the Clostridium botulinum type-E
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT gene.";
RL EUR. J. BIOCHEM. 204:657-667(1992).
RN [3]
RP SEQUENCE OF 1-251 FROM N.A.
RX MEDLINE; 90264400.
RA BINZ T., KURAZONO H., WILLE M., FREVERT J., WERNARS K., NIEMANN H.;
RT "The complete sequence of botulinum neurotoxin type A and comparison
RT with other clostridial neurotoxins.";
RL J. BIOL. CHEM. 265:9153-9158(1990).
RN [4]
RP SEQUENCE OF 1-13.
RX MEDLINE; 85197963.
RA SCHMIDT J.J., SATHYAMOORTHY V., DASGUPTA B.R.;
RT "Partial amino acid sequences of botulinum neurotoxins types B and
RT E.";
RL ARCH. BIOCHEM. BIOPHYS. 238:544-548(1985).
RN [5]
RP SEQUENCE OF 419-426.
RX MEDLINE; 90344918.
RA GIMENEZ J.A., DASGUPTA B.R.;
RT "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT reveals the site trypsin nicks and homology with tetanus
RL neurotoxin.";
RL BIOCHIMIE 72:213-217(1990).
RN [6]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 94063091.
RA SCHIAVO G., SANTUCCI A., DASGUPTA B.R., MEHTA P.P., JONTES J.,
RA BENFENATI F., WILSON M.C., MONTECUCCHI C.;
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT COOH-terminal peptide bonds.";
RL FEBS LETT. 335:99-103(1993).
RN [7]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE; 94124495.
RA BINZ T., BLASI J., YAMASAKI S., BAUMEISTER A., LINK E., SUDHOF T.C.,
RA JAHN R., NIEMANN H.;
RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";
RL J. BIOL. CHEM. 269:1617-1620(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

```
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-1-ILE-
CC 181 BOND IN SNAP-25.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62089; G40394; -.
DR EMBL: X62683; G40398; -.
DR PIR: A60027; A60027.
DR PIR: B35294; B35294.
DR PIR: JH0257; JH0257.
DR PIR: S08575; S08575.
DR PIR: S18111; S18111.
DR PIR: S21178; S21178.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PFAM: PF00099; zn-protease; 1.
DR HSSP: P04958; 1AF9.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT INIT_MET 0
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT METAL 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT ACT_SITE 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 212 212 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 176 176 R -> G (IN REF. 2).
FT CONFLICT 197 197 C -> S (IN REF. 2 AND 3).
FT CONFLICT 339 339 R -> A (IN REF. 2).
FT CONFLICT 772 772 I -> L (IN REF. 2).
FT CONFLICT 962 963 FE -> LQ (IN REF. 2).
FT CONFLICT 966 966 R -> A (IN REF. 2).
FT CONFLICT 1194 1194 N -> NN (IN REF. 2).
FT CONFLICT 1250 1250 AA; 143712 MW; 4BE9332D CRC32;
SQ SEQUENCE

Query Match 55.7%; Score 88; DB 1; Length 1250;
Best Local Similarity 50.0%; Pred. No. 1.11e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db 911 YKNFSISFWVRIPN 924
QY 1 FNNFTVSWLRVPK 14

RESULT 7
ID BXE_CLOBU STANDARD; PRT; 1250 AA.
AC P30995;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE E PRECURSOR (EC 3.1.24.69) (BONT/E)
DE (BONTOLYSIN E).
OS CLOSTRIDIUM BUTYRICUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN-ATCC 43181, AND ATCC 43755;
RX MEDLINE; 92181428.
RA POUTLET S., HAUSER D., QUANZ M., NIEMANN H., POPOFF M.R.;
RT "Sequences of the botulinum neurotoxin E derived from Clostridium
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT ATCC 43181 and ATCC 43755).";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:107-113(1992).
RN [2]
RP SEQUENCE OF 1-251 FROM N.A.
RC STRAIN-BL6340;
RX MEDLINE; 91237316.
RA FUJII N., KIMURA K., MURAKAMI T., INDOH T., TSUZUKI K.,
RA YOKOSAWA N., YASHIKI T., OGUMA K.;
RT "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT type E toxin gene from Clostridium butyricum strain BL6340.";
RL J. GEN. MICROBIOL. 137:519-525(1991).
RN [3]
RP SEQUENCE OF 1-48.
RC STRAIN-5262;
RA GIMENEZ J., FOLEY J., DASGUPTA B.R.;
RT "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT partial sequence and comparison.";
RL FASEB J. 2:A1750-A1750(1988).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62088; G40380; -.
DR EMBL: X53180; G40408; -.
DR PIR: JH0256; JH0256.
DR PIR: S16145; S16145.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PFAM: PF00099; zn-protease; 1.
DR HSSP: P04958; 1AF9.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT INIT_MET 0
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT METAL 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT ACT_SITE 211 211 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 212 212 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 411 425 INTERCHAIN (PROBABLE).
FT CONFLICT 229 229 K -> M (IN REF. 2).
FT CONFLICT 1250 1250 AA; 143265 MW; BC046B39 CRC32;
SQ SEQUENCE

Query Match 55.7%; Score 88; DB 1; Length 1250;
Best Local Similarity 50.0%; Pred. No. 1.11e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db 911 YKNFSISFWVRIPN 924
QY 1 FNNFTVSWLRVPK 14
```

```
RESULT 8
ID VP2_AHSV6 STANDARD; PRT; 1051 AA.
AC 071024;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP2.
GN S2 OR I2.
OS AFRICAN HORSE SICKNESS VIRUS 6 (AHSV-6) (AFRICAN HORSE SICKNESS VIRUS
OS (SEROTYPE 6)).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ORBIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98278331.
RA WILLIAMS C.F., INOUE T., LUCUS A.-M., ZANOTTO P., ROY P.;
RT "The complete sequence of four major structural proteins of African
RT horse sickness virus serotype 6: evolutionary relationships within
RT and between the orbiviruses."
RL VIRUS RES. 53:53-73(1998).
CC -1- FUNCTION: THE VP2 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP5)
CC WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID. IT IS THE
CC MAJOR TARGET OF THE HOST IMMUNOGENIC RESPONSE.
CC -1- SIMILARITY: BELONGS TO THE THE REOVIRUSES VP2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF021235; G3168848; -
KW COAT PROTEIN.
SQ SEQUENCE 1051 AA; 122326 MW; 05DB1A29 CRC32;
Db 636 FSKRFVSYWYRVEKITTKHLE 656
QY 1 FNNFTVSFWLRVPKVSASHLE 21
Query Match 55.1%; Score 87; DB 1; Length 1051;
Best Local Similarity 47.6%; Pred. No. 1.76e-04;
Matches 10; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
RESULT 9
ID VNSS_TSWVL STANDARD; PRT; 464 AA.
AC P26002;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN NS-S.
GN NSS.
OS TOMATO SPOTTED WILT VIRUS (STRAIN BRAZILIAN BR-01) (TSWV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90264829.
RA DE HAAN P., WAGEMAKERS L., PETERS D., GOLDBACH R.;
RT "The S RNA segment of tomato spotted wilt virus has an ambisense
RT character."
RL J. GEN. VIROL. 71:1001-1007(1990).
CC -1- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL; D00645; D1000995; -
DR PIR; JQ0547; MNAVWC.
KW NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 464 AA; 52448 MW; BC073BFC CRC32;
Db 239 SHEKLSLWLRVPKV 252
QY 2 NNFVTSFWLRVPKV 15
Query Match 53.2%; Score 84; DB 1; Length 464;
Best Local Similarity 64.3%; Pred. No. 6.88e-04;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
RESULT 10
ID VNSS_TSWVL STANDARD; PRT; 467 AA.
AC P26003;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DE 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE NONSTRUCTURAL PROTEIN NS-S.
GN NSS.
OS TOMATO SPOTTED WILT VIRUS (STRAIN BULGARIAN I3) (TSWV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91132150.
RA MAISS E., IVANOVA L., BREYEL E., ADAM G.;
RT "Cloning and sequencing of the S RNA from a Bulgarian isolate of
RT tomato spotted wilt virus."
RL J. GEN. VIROL. 72:461-464(1991).
CC -1- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D13926; G222683; -
DR PIR; JQ0954; MNAVWL.
KW NONSTRUCTURAL PROTEIN.
SQ SEQUENCE 467 AA; 52413 MW; F2EF7412 CRC32;
Db 243 SHEKLSLWLRVPKV 256
QY 2 NNFVTSFWLRVPKV 15
Query Match 53.2%; Score 84; DB 1; Length 467;
Best Local Similarity 64.3%; Pred. No. 6.88e-04;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
RESULT 11
ID BXCL_CLOBO STANDARD; PRT; 1290 AA.
AC P18640;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DE 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE C1 PRECURSOR (EC 3.4.24.69) (BONT/C1)
DE (BONTXILYSIN C1).
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90370487.
RA HAUSER D., EKLUND M.W., KURAZONA H., BINZ T., NIEMANN H., GILL D.M.,
RA BOQUEI P., POPOFF M.R.;
RT "Nucleotide sequence of Clostridium botulinum C1 neurotoxin.";
```

RL NUCLEIC ACIDS RES. 18:4924-4924(1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=TYPE C STOCKHOLM / C-ST;

RX MEDLINE; 91024998.

RA KIMURA K., FUJII N., TSUZUKI K., MURAKAMI T., INDOH T.,

RA YOKOSAWA N., TAKESHI K., SYUTO B., OGUMA K.;

RT "The complete nucleotide sequence of the gene coding for botulinum

RT type C1 toxin in the C-ST phage genome.";

RL BIOCHEM. BIOPHYS. RES. COMMUN. 171:1304-1311(1990).

RN [3]

RP SEQUENCE OF 2-25.

RC STRAIN=TYPE C STOCKHOLM / C-ST;

RX MEDLINE; 88153072.

RA TSUZUKI K., YOKOSAWA N., SYUTO B., OHISHI I., FUJII N., KIMURA K.,

RA OGUMA K.;

RT "Establishment of a monoclonal antibody recognizing an antigenic site

RT common to Clostridium botulinum type B, C1, D, and E toxins and

RT tetanus toxin.";

RL INFECT. IMMUN. 56:898-902(1988).

RN [4]

RP IDENTIFICATION OF SUBSTRATE.

RX MEDLINE; 94038966.

RA BLASI J., CHAPMAN E.R., YAMASAKI S., BINZ T., NIEMANN H., JAHN R.;

RT "Botulinum neurotoxin C1 blocks neurotransmitter release by means of

RT cleaving HPC-1/syntaxin.";

RL EMBO J. 12:4821-4828(1993).

CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE THAT CLEAVES SYNTAXIN.

CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL

CC ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE

CC CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM

CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -1- BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C STRAIN OF

CC CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE BACTERIOPHAGE.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC

CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X66433; G509275; -.

DR EMBL; X72793; G516175; -.

DR EMBL; X53751; G14906; -.

DR EMBL; D90210; G217781; -.

DR EMBL; X62383; G40390; -.

DR PIR; S11291; S11291.

DR PIR; A35396; A35396.

DR PIR; A43503; A43503.

DR PROSITE; PS00142; ZINC_PROTEASE; 1.

DR PRAM; PF00099; zn-protease; 1.

KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.

FT INTL_MET 0

FT CHAIN 1 448

FT CHAIN 449 1290

FT METAL 228

FT ACT_SITE 229

FT METAL 232

FT DISULFID 436 452

FT CONFLICT 84

BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.

BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.

ZINC (CATALYTIC) (BY SIMILARITY).

BY SIMILARITY.

ZINC (CATALYTIC) (BY SIMILARITY).

INTERCHAIN (PROBABLE).

P -> T (IN REF. 2).

SQ SEQUENCE 1290 AA; 148734 MW; DOBCB190 CRC32;

Query Match

Best Local Similarity 42.9%; Pred. No. 2.63e-03;

Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 934 YESFISFWIRINK 947

QY 1 FNNFTVSWLRVPK 14

RESULT 12

ID VNSS_INSVN STANDARD; PRT; 449 AA.

AC Q01811;

DT 01-JUL-1993 (REL. 26, CREATED)

DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)

DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)

DE NONSTRUCTURAL PROTEIN NS-S.

GN NSS.

OS IMPATIENS NECROTIC SPOT VIRUS (STRAIN NL-07) (INSV).

OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE; 92331780.

RA DE HAAN P., DE AVILA A.C., KORMELINK R., WESTERBROEK A.,

RA GIELEN J.J., PETERS D., GOLDBACH R.;

RT "The nucleotide sequence of the S RNA of Impatiens necrotic spot

RT virus, a novel tospovirus.";

RL FEBS LETT. 306:27-32(1992).

CC -1- FUNCTION: FORMS FILAMENTOUS INCLUSION BODIES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X66972; G60489; -.

DR PIR; S23158; S23158.

KW NONSTRUCTURAL PROTEIN.

SQ SEQUENCE 449 AA; 51197 MW; 6A87666F CRC32;

Query Match

Best Local Similarity 48.7%; Score 77; DB 1; Length 449;

Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 244 FKISLWMRIPIK 255

QY 4 FTVSWLRVPK 15

RESULT 13

ID BXCN_CLOBO STANDARD; PRT; 1196 AA.

AC P46081;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)

DE BOTULINUM NEUROTOXIN TYPE C1, NONTOXIC COMPONENT.

OS CLOSTRIDIUM BOTULINUM.

OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;

OC CLOSTRIDIUM.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TYPE C STOCKHOLM / C-ST;

RX MEDLINE; 92231894.

RA TSUZUKI K., KIMURA K., FUJII N., YOKOSAWA N., OGUMA K.;

RT "The complete nucleotide sequence of the gene coding for the

RT nontoxic-nonhemagglutinin component of Clostridium botulinum type C

RT progenitor toxin.";

RL BIOCHEM. BIOPHYS. RES. COMMUN. 183:1273-1279(1992).

CC -1- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN


```

CC      TOXICITY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X62389; G40389; -.
KW      NEUROTOXIN.
SQ      SEQUENCE 1196 AA; 138740 MW; E5C11933 CRC32;

Query Match      48.1%; Score 76; DB 1; Length 1196;
Best Local Similarity 80.0%; Pred. No. 2.32e-02;
Matches      8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db      899 NNETVSWFLR 908
QY      2 NNETVSWFLR 11

RESULT 14
ID      BXD_CLOBO      STANDARD;      PRT; 1276 AA.
AC      P19321;
DT      01-NOV-1990 (REL. 16, CREATED)
DT      01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT      15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE      BOTULINUM NEUROTOXIN TYPE D PRECURSOR (EC 3.4.24.69) (BONT/D)
GN      BONTOLIXIN D).
OS      CLOSTRIDIUM BOTULINUM.
OC      BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC      CLOSTRIDIUM.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BYD/-3;
RX      MEDLINE; 91016853.
RA      BINZ T., KURAZONO H., POPOFF M.R., EKLUND M.W., SAKAGUCHI G.,
RA      KOZAKI S., KRIEGLSTEIN K., HENSCHEN A., GILL D.M., NIEMANN H.;
RT      "Nucleotide sequence of the gene encoding clostridium botulinum
RT      neurotoxin type D.";
RL      NUCLEIC ACIDS RES. 18:5556-5556(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CB16;
RX      MEDLINE; 93042276.
RA      SUNAGAWA H., OHYAMA T., WATANABE T., INOUE K.;
RT      "The complete amino acid sequence of the clostridium botulinum type D
RT      neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT      phage d-16 phi genome.";
RL      J. VET. MED. SCI. 54:905-913(1992).
RN      [3]
RP      PARTIAL SEQUENCE.
RC      STRAIN=D-SA, AND D-1873;
RX      MEDLINE; 89339741.
RA      MORITISHI K., SYUTO B., KUBO S., OGUMA K.;
RT      "Molecular diversity of neurotoxins from clostridium botulinum type D
RT      strains.";
RL      INFECT. IMMUN. 57:2886-2891(1989).
RN      [4]
RP      IDENTIFICATION OF SUBSTRATE.
RC      MEDLINE; 94230352.
RA      YAMASAKI S., BAUMEISTER A., BINZ T., BLASI J., LINK E., CORNILLE F.,
RA      ROQUES B., FYKSE E.M., SUEDEHOF T.C., JAHN R., NIEMANN H.;
RT      "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT      F botulin neurotoxins and tetanus toxin.";
RL      J. BIOL. CHEM. 269:12764-12772(1994).
RN      [5]
RP      FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC      RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC      AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC      WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

```

```

CC      INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC      ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-1-LEU-61 BOND OF
CC      SYNAPTOBREVIN-1 AND -2.
CC      -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC      A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL
CC      ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE
CC      CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC      -1- SUBCELLULAR LOCATION: SECRETED.
CC      -1- THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM
CC      NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC      -1- BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D STRAIN OF
CC      CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE BACTERIOPHAGE.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
CC      METALLOPROTEASE); ALSO KNOWN AS THE TEIANUS/BOTULINUM NEUROTOXIN
CC      SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: X54254; G40396; -.
DR      EMBL: S49407; G260239; -.
DR      PIR: S11455; S11455.
DR      PROSITE: PS00142; ZINC_PROTEASE; 1.
DR      PFAM: PF00099; zn-protease; 1.
KW      NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
FT      CHAIN 1 442
FT      CHAIN 2 442
FT      METAL 443 1276
FT      METAL 229 229
FT      ACT_SITE 230 230
FT      METAL 233 233
FT      DISULFID 437 450
FT      VARIANT 15 16
FT      VARIANT 17 18
FT      VARIANT 452 452
FT      VARIANT 457 457
FT      VARIANT 457 457
FT      VARIANT 457 457
FT      VARIANT 462 462
FT      VARIANT 489 489
FT      VARIANT 644 644
FT      VARIANT 1122 1122
SQ      SEQUENCE 1276 AA; 146871 MW; 02D7FECC CRC32;

Query Match      45.6%; Score 72; DB 1; Length 1276;
Best Local Similarity 40.9%; Pred. No. 1.25e-01;
Matches      9; Conservative 9; Mismatches 3; Indels 1; Gaps 1;

Db      928 YENSSVSFWIKSKDLTNSHNE 949
QY      1 FNNFTVSWFLRYPK-VSASHLE 21

RESULT 15
ID      BXBEN_CLOBO      STANDARD;      PRT; 1162 AA.
AC      P46082;
DT      01-NOV-1995 (REL. 32, CREATED)
DT      01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT      01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE      BOTULINUM NEUROTOXIN TYPE E, NONTOXIC COMPONENT.
GN      ENT-120.
OS      CLOSTRIDIUM BOTULINUM.
OC      BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC      CLOSTRIDIUM.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=MASHIKE;
RX      MEDLINE; 93195515.
RA      FUJII N., KIMURA K., YOKOSAWA N., YASHIKI T., TSUZUKI K., OGUMA K.;
RT      "The complete nucleotide sequence of the gene encoding the nontoxic

```

```

RT component of Clostridium botulinum type E progenitor toxin.";
RL J. GEN. MICROBIOL. 139:79-86(1993).
CC -1- FUNCTION: THE NONTOXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D12697; G285708; -.
KW NEUROTOXIN.
SQ SEQUENCE 1162 AA; 136856 MW; F650831D CRC32;

```

```

Query Match 44.3%; Score 70; DB 1; Length 1162;
Best Local Similarity 70.0%; Pred. No. 2.84e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 864 NNESTYFWLR 873
QY 2 NNFTVSEWLR 11

```

Search completed: Wed Aug 4 15:25:45 1999
Job time : 5 secs.

[M] [E] [R] [E] [L] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:26:03 1999; MasPar time 7.37 Seconds

Tabular output not generated. 155.440 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 179066 segs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 29.394; Variance 45.917; scale 0.640

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	95	60.1	1268	2	Q45851	NEUROTOXIN TYPE F.	1.58e-05
2	93	58.9	361	2	Q45848	BOTULINUM NEUROTOXIN T	3.93e-05
3	93	58.9	361	2	Q45846	BOTULINUM NEUROTOXIN T	3.93e-05
4	93	58.9	1278	2	Q57236	BONT/F (BOTULINUM NEUR	3.93e-05
5	93	58.9	1291	2	Q08077	BOTULINUM NEUROTOXIN T	3.93e-05
6	88	55.7	367	2	Q45862	BOTULINUM NEUROTOXIN T	3.68e-04
7	88	55.7	367	2	Q45861	BOTULINUM NEUROTOXIN T	3.68e-04
8	85	53.8	1296	2	Q45894	BOTULINUM NEUROTOXIN T	1.38e-03
9	84	53.2	467	14	Q37369	NON-STRUCTURAL PROTEIN	2.13e-03
10	84	53.2	467	14	Q37367	NON-STRUCTURAL PROTEIN	2.13e-03
11	82	51.9	467	14	Q88900	NON-STRUCTURAL PROTEIN	5.03e-03
12	81	51.3	1285	5	Q45967	NEUROTOXIN CONSISTING	7.72e-03
13	79	50.0	322	5	P91551	COSMID ZC239.	1.80e-02
14	76	48.1	1194	2	Q45916	138KDA PROTEIN ASSOCIA	6.29e-02
15	76	48.1	1196	2	Q53550	PROGENITOR TOXIN L NON	6.29e-02
16	76	48.1	1196	9	Q38197	ANTP-139.	6.29e-02
17	75	47.5	175	5	P91554	COSMID ZC239.	9.49e-02
18	75	47.5	558	5	Q76933	PENTAXIN-LIKE PROTEIN	9.49e-02
19	72	45.6	406	2	Q53946	HYPOTHETICAL 44.2 KD P	3.20e-01
20	72	45.6	1280	2	Q45849	NEUROTOXIN.	3.20e-01

21	69	43.7	637	4	Q14744	SKBLHS.	1.05e+00
22	69	43.7	1197	2	Q33871	NONTOXIC-NONHEMAGGLUTI	1.05e+00
23	69	43.7	1197	2	Q45888	NONTOXIC-HEMAGGLUTININ	1.05e+00
24	69	43.7	1197	2	P71117	NONTOXIC-NONHEMAGGLUTI	1.05e+00
25	69	43.7	1198	2	Q06018	PROGENITOR TOXIN COMPL	1.05e+00
26	68	43.0	699	4	Q14978	NUCLEOLAR PHOSPHOPROTE	1.55e+00
27	68	43.0	707	4	Q15030	ORE (FRAGMENT)	1.55e+00
28	67	42.4	152	2	P76263	FROM BASES 1896340 TO	2.28e+00
29	66	41.8	244	2	Q05883	HYPOTHETICAL 28.0 KD P	3.33e+00
30	66	41.8	1159	2	P71109	NONTOXIC-NONHEMAGGLUTI	3.33e+00
31	66	41.8	1160	2	Q87710	NTNH PROTEIN.	3.33e+00
32	66	41.8	1161	2	Q45891	NTNH PROTEIN.	3.33e+00
33	66	41.8	1161	2	Q69276	NONTOXIC-NONHEMAGGLUT	3.33e+00
34	66	41.8	1165	2	Q45887	BOTULINUM NEUROTOXIN T	3.33e+00
35	66	41.8	1165	2	Q45844	NEUROTOXIN COMPLEX M N	3.33e+00
36	66	41.8	1193	2	Q45914	TYPE A PROGENITOR TOXI	3.33e+00
37	66	41.8	1193	2	P71107	A NTNH AND BONT GENES.	3.33e+00
38	66	41.8	1193	2	P71108	NONTOXIC-NONHEMAGGLUTI	3.33e+00
39	66	41.8	1193	2	Q45880	NTNHA.	3.33e+00
40	66	41.8	1198	2	Q45893	NTNH PROTEIN.	3.33e+00
41	66	41.8	1198	2	Q69277	NONTOXIC-NONHEMAGGLUT	3.33e+00
42	65	41.1	203	2	Q06370	REPRESSOR PROTEIN OF S	4.87e+00
43	65	41.1	510	1	Q51733	BETA-MANNOSIDASE.	4.87e+00
44	65	41.1	786	8	Q34832	INTRON ORF.	4.87e+00
45	65	41.1	834	14	Q41921	NONSTRUCTURAL POLYPEPT	4.87e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY;	PRT;	1268 AA.
1	Q45851			
AC	Q45851;			
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)			
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	NEUROTOXIN TYPE F.			
GN	BONT /F.			
OS	CLOSTRIDIUM BARATI.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC	CLOSTRIDIUM.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 93252228.			
RA	THOMPSON D.E., HUTSON R.A., EAST A.K., ALLAWAY D., COLLINS M.D.,			
RA	RICHARDSON P.T.;			
RT	"Nucleotide sequence of the gene coding for Clostridium barati type F			
RT	neurotoxin: comparison with other clostridial neurotoxins.";			
RL	EMMS MICROBIOL. LETT. 108:175-182(1993).			
DR	EMBL; X68262; G49139; -			
DR	PFAM; PF00099; zn-protease; 1.			
KW	NEUROTOXIN.			
SO	SEQUENCE	1268 AA;	145512 MW;	37B7AA07 CRC32;
Query Match				
Best Local Similarity 60.1%; Score 95; DB 2; Length 1268;				
Matches 9; Conservativity 4; Mismatches 1; Indels 0; Gaps 0;				
Db	922 YONFSVSFWVRIPK 935			
QY	1 FNNFTVSFWLRVPK 14			
RESULT 2				
ID	Q45848	PRELIMINARY;	PRT;	361 AA.
AC	Q45848;			
DT	01-NOV-1996 (TREMBLREL. 01, CREATED)			
DT	01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).			
GN	BONT/B.			
OS	CLOSTRIDIUM BOTULINUM.			
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;			
OC	CLOSTRIDIUM.			

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B, NON-PROTEOLYTIC EKLUND 2B (COLWORTH 229);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X70819; G407781; -.
KW NEUROTOXIN.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42131 MW; 8FBF998C CRC32;

Query Match
Best Local Similarity 58.9%; Score 93; DB 2; Length 361;
Matches 9; Conservativity 4; Mismatches 1; Indels 0; Gaps 0;

Db 290 FLDFSVSFWIRPK 303
QY 1 FNNFTVSEWLRVPK 14

RESULT 3
ID Q45846 PRELIMINARY; PRT; 361 AA.
AC Q45846;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN BONT/B.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X70814; G407779; -.
KW NEUROTOXIN.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 42175 MW; 3863DE86 CRC32;

Query Match
Best Local Similarity 58.9%; Score 93; DB 2; Length 361;
Matches 9; Conservativity 4; Mismatches 1; Indels 0; Gaps 0;

Db 290 FLDFSVSFWIRPK 303
QY 1 FNNFTVSEWLRVPK 14

RESULT 4
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BONT/F (BOTULINUM NEUROTOXIN TYPE F).
GN BONT/F.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA HUTSON R.A., COLLINS M.D.;
RT SUBMITTED (AUG-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA ELMORE M.J., BODSWORTH N.J., WHELAN S.M., MINTON N.P.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=TYPE F, PROTEOLYTIC F LANGE LAND (NCTC 1028);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X81714; G971349; -.
DR EMBL; L35496; G529984; -.
DR EMBL; X70821; G407793; -.
DR PFAM; PF00099; zn-protease; 1.
KW NEUROTOXIN.
SQ SEQUENCE 1278 AA; 147073 MW; 56C58E6F CRC32;

Query Match
Best Local Similarity 58.9%; Score 93; DB 2; Length 1278;
Matches 8; Conservativity 5; Mismatches 1; Indels 0; Gaps 0;

Db 931 YQNSISFWIRPK 944
QY 1 FNNFTVSEWLRVPK 14

RESULT 5
ID Q08077 PRELIMINARY; PRT; 1291 AA.
AC Q08077;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE B (EC 3.4.24.-) (BONT/B).
GN BONT/B.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EKLUND 17B ATCC25765;
RX MEDLINE; 94122659.
RA HUTSON R.A., COLLINS M.D., EAST A.K., THOMPSON D.E.;
RT "Nucleotide sequence of the gene coding for non-proteolytic
Clostridium botulinum type B neurotoxin: comparison with other
Clostridial neurotoxins.";
RL CURR. MICROBIOL. 28:101-110(1994).
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
RELEASE.
CC IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND
MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS.
CC IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
ENDOPEPTIDASE THAT CLEAVES SYNAPTOSOMAL-2.
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN
(L) AND A HEAVY CHAIN (H).
CC THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-
AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND
TOXIN BINDING, RESPECTIVELY.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC -1- SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH
TETANUS NEUROTOXIN.
CC -1- SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE
REGION.
DR EMBL; X71343; G296149; -.
DR PFAM; PF00099; zn-protease; 1.
KW NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC.
SQ SEQUENCE 1291 AA; 150513 MW; 5210A9B7 CRC32;

Query Match
Best Local Similarity 58.9%; Score 93; DB 2; Length 1291;

```

Best Local Similarity 64.3%; Pred. No. 3.93e-05;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 923 FLDFSVSFWIRPK 936

QY 1 FNNFTVSFWLRVPK 14

```
RESULT 6
ID 045862 PRELIMINARY; PRT; 367 AA.
AC 045862;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN BONT/E.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE E, HAZEN 36208 (ATCC 9564);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X70815; G407787; -.
KW NEUROTOXIN.
ET NON_TER 1
FT NON_TER 367
SQ SEQUENCE 367 AA; 42854 MW; 95DDBA66 CRC32;
```

Query Match 55.7%; Score 88; DB 2; Length 367;
Best Local Similarity 50.0%; Pred. No. 3.68e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

Db 297 YKNFSISFWVRIPN 310

QY 1 FNNFTVSFWLRVPK 14

```
RESULT 7
ID 045861 PRELIMINARY; PRT; 367 AA.
AC 045861;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN BONT/E.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE E, VH (DOLMAN);
RX MEDLINE; 94013372.
RA CAMPBELL K., EAST A.K., COLLINS M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. CLIN. MICROBIOL. 31:2255-2262(1993).
DR EMBL; X70818; G407785; -.
KW NEUROTOXIN.
ET NON_TER 1
FT NON_TER 367
SQ SEQUENCE 367 AA; 42902 MW; 695DD505 CRC32;
```

Query Match 55.7%; Score 88; DB 2; Length 367;
Best Local Similarity 50.0%; Pred. No. 3.68e-04;
Matches 7; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 FNNFTVSFWLRVPK 14

```
RESULT 8
ID 045894 PRELIMINARY; PRT; 1296 AA.
AC 045894; P77780;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE BOTULINUM NEUROTOXIN TYPE A.
GN BONT.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KYOTO-F;
RX MEDLINE; 94143603.
RA WILLEMS A., EAST A.K., LAWSON P.A., COLLINS M.D.;
RT "Sequence of the gene coding for the neurotoxin of Clostridium
RT botulinum type A associated with infant botulism: comparison with
RT other clostridial neurotoxins.";
RL RES. MICROBIOL. 144:547-556(1993).
RN [2]
RP SEQUENCE OF 1-65 FROM N.A.
RC STRAIN=62A;
RX MEDLINE; 97016817.
RA EAST A.K., BRANDARI M., STACEY J.M., CAMPBELL K.D., COLLINS M.D.;
RT "Organization and phylogenetic interrelationships of genes encoding
RT components of the botulinum toxin complex in proteolytic Clostridium
RT botulinum types A, B, and F: evidence of chimeric sequences in the
RT gene encoding the nontoxic nonhemagglutinin component.";
RL INT. J. SYST. BACTERIOL. 46:1105-1112(1996).
DR EMBL; X73423; G507071; -.
DR EMBL; X92973; E212189; -.
DR EMBL; X87974; E184381; -.
KW NEUROTOXIN.
SQ SEQUENCE 1296 AA; 149410 MW; 24AF86B1 CRC32;
```

Query Match 53.8%; Score 85; DB 2; Length 1296;
Best Local Similarity 50.0%; Pred. No. 1.38e-03;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 938 YENFSTFWIKIPK 951

QY 1 FNNFTVSFWLRVPK 14

```
RESULT 9
ID 037369 PRELIMINARY; PRT; 467 AA.
AC 037369;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE NON-STRUCTURAL PROTEIN.
GN NSS.
OS TOMATO SPOTTED WILT VIRUS (TSWV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
OC VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TSWV-D;
RX MEDLINE; 98240946.
RA QIU W.-P., GESKE S.M., HICKEY C.M., MOYER J.W.;
RT "Tomato spotted wilt tospovirus genome reassortment and genome
RT segment-specific adaptation.";
RL VIROLOGY 244:186-194(1998).
DR EMBL; AF020660; G2425149; -.
SQ SEQUENCE 467 AA; 52587 MW; 1D153724 CRC32;
```

Query Match 53.2%; Score 84; DB 14; Length 467;
Best Local Similarity 64.3%; Pred. No. 2.13e-03;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

```
Db      243 SHEKLSIMLRVPKV 256
      :|:|:|:|:|:|
QY      2 NNFTVSFWLRVPKV 15

RESULT  10
ID      037367          PRELIMINARY;      PRT;      467 AA.
AC      037367;
DT      01-JAN-1998 (TREMBLREL. 05, CREATED)
DT      01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      NON-STRUCTURAL PROTEIN.
GN      NSS.
OS      TOMATO SPOTTED WILT VIRUS (TSWV).
OC      VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE; TOSPOVIRUS.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TSW-10;
RX      MEDLINE; 98240946.
RA      QIU W.-P., GESKE S.M., HECKEY C.M., MOYER J.W.;
RT      "Tomato spotted wilt tospovirus genome reassortment and genome
RL      segment-specific adaptation.";
RL      VIROLOGY 244:186-194(1998).
DR      EMBL; AF020659; G2425152; -.
SQ      SEQUENCE 467 AA; 52505 MW; B3A4243F CRC32;

Query Match      53.2%; Score 84; DB 14; Length 467;
Best Local Similarity 64.3%; Pred. No. 2.13e-03;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db      243 SHEKLSIMLRVPKV 256
      :|:|:|:|:|:|
QY      2 NNFTVSFWLRVPKV 15

RESULT  11
ID      088900          PRELIMINARY;      PRT;      467 AA.
AC      088900;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT      01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE      NONSTRUCTURAL PROTEIN.
OS      UNKNOWN.
OC      VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; BUNYAVIRIDAE.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=TSW-B;
RA      PANG S.Z., SLIGHTOM J.L., GONSALVES D.;
RL      PHYTOPATHOLOGY 83:728-733(1993).
DR      EMBL; L12048; G335275; -.
KW      NONSTRUCTURAL PROTEIN.
SQ      SEQUENCE 467 AA; 52565 MW; BD32D7C7 CRC32;

Query Match      51.9%; Score 82; DB 14; Length 467;
Best Local Similarity 57.1%; Pred. No. 5.03e-03;
Matches 8; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db      243 SHEKLSIMLRIPKV 256
      :|:|:|:|:|:|
QY      2 NNFTVSFWLRVPKV 15

RESULT  12
ID      045967          PRELIMINARY;      PRT;      1285 AA.
AC      045967;
DT      01-NOV-1996 (TREMBLREL. 01, CREATED)
DT      01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT      01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE      NEUROTOXIN CONSISTING OF BOTULINUM NEUROTOXIN D AND C1.
OS      CLOSTRIDIUM BOTULINUM.
OC      BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
KW      CLOSTRIDIUM.
RN      [1]
```

```
RP      SEQUENCE FROM N.A.
RC      STRAIN=TYPE D, SOUTH AFRICAN;
RX      MEDLINE; 96283801.
RA      MORIISHI K., KOURA M., ABE N., FUJII N., FUJINAGA Y., INOUE K.,
RA      OGUMAD K.;
RT      "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT      types C and D organisms.";
RL      BIOCHIM. BIOPHYS. ACTA 1307:123-126(1996).
DR      EMBL; D38442; D1008057; -.
DR      PFAM; PF00099; zn-protease; 1.
KW      NEUROTOXIN.
SQ      SEQUENCE 1285 AA; 147366 MW; 381133E1 CRC32;

Query Match      51.3%; Score 81; DB 2; Length 1285;
Best Local Similarity 42.9%; Pred. No. 7.72e-03;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db      931 YESSESFWIRINK 944
      :|:|:|:|:|:|
QY      1 FNNFTVSFWLRVPK 14

RESULT  13
ID      P91551          PRELIMINARY;      PRT;      322 AA.
AC      P91551;
DT      01-MAY-1997 (TREMBLREL. 03, CREATED)
DT      01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT      01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE      COSMID ZC239.
GN      ZC239.8.
OS      CAENORHABDITIS ELEGANS.
OC      EUKARYOTA; METAZOA; NEMATODA; SECERNENTEAE; RHABDITIA; RHABDITIDA;
OC      RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2;
RX      MEDLINE; 94150718.
RA      WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA      BONEFIELD J., BORTON J., CONNELL M., COPSEY I., COOPER C., COULSON A.,
RA      CRAFTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA      GARDNER A., DEAR P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA      JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA      LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA      PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA      SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STILSTON J.,
RA      THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA      WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT      "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT      elegans.";
RT      NATURE 368:32-38(1994).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2;
RA      WU X., KRAMER J.;
RL      SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN-BRISTOL N2;
RA      WATERSTON R.;
RL      SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR      EMBL; U80842; G1707209; -.
SQ      SEQUENCE 322 AA; 36968 MW; B4E0F90C CRC32;

Query Match      50.0%; Score 79; DB 5; Length 322;
Best Local Similarity 70.0%; Pred. No. 1.80e-02;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db      4 DFTVTFWLRI 13
      :|:|:|:|:|:|
QY      3 NETVSFWLRV 12

RESULT  14
```

ID Q45916 PRELIMINARY; PRT; 1194 AA.
AC Q45916;
DI 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DI 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE 138KDA PROTEIN ASSOCIATED WITH BONT /C1-HAEMAGGLUTININ COMPLEX.
GN CHN-138.
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA HAUSER D.F., EKLUND M.W., POPOFF M.R.;
RL SUBMITTED (MAY-1992) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X66433; G509274;
SQ SEQUENCE 1194 AA; 138509 MW; 0192832A CRC32;

Query Match 48.1%; Score 76; DB 2; Length 1194;
Best Local Similarity 80.0%; Pred. No. 6.29e-02;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 899 NNETVCFWLR 908
QY 2 NNETVCFWLR 11

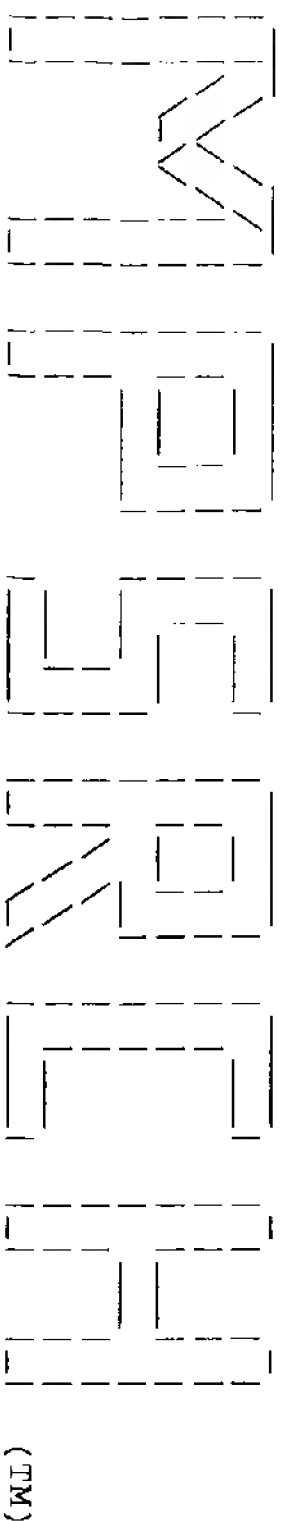
RESULT 15
ID Q53550 PRELIMINARY; PRT; 1196 AA.
AC Q53550;
DI 01-NOV-1996 (TREMBLREL. 01, CREATED)
DI 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DI 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PROGENITOR TOXIN L NONTOXIC-NONHEMAGGLUTININ COMPONENT (FRAGMENT).
OS CLOSTRIDIUM BOTULINUM.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;
OC CLOSTRIDIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96025415.
RA OHYAMA T., WATANABE T., FUJINAGA Y., INOUE K., SUNAGAWA H., FUJII N.,
RA INOUE K., OGUMA K.;
RT "Characterization of nontoxic-nonhemagglutinin component of the two
RT types of progenitor toxin (M and L) produced by Clostridium botulinum
RT type D CB-16."
RL MICROBIOL. IMMUNOL. 39:457-465(1995).
DR EMBL; S80809; G1311691; -.
FT NON_TER 1196 1196
SQ SEQUENCE 1196 AA; 138717 MW; B534D1B0 CRC32;

Query Match 48.1%; Score 76; DB 2; Length 1196;
Best Local Similarity 80.0%; Pred. No. 6.29e-02;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 899 NNETVCFWLR 908
QY 2 NNETVCFWLR 11

Search completed: Wed Aug 4 15:26:13 1999
Job time : 10 secs.

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:48:12 1999; MasPar time 3.68 Seconds

Tabular output not generated. 121.425 Million cell updates/sec

Title: >US-09-049-847-2

Description: (1-21) from US09049847.pep

Perfect Score: 158

Sequence: 1 FNNFTVSFWLRVPKVSASHLE 21

Scoring table: PAM 150

Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 21.709; Variance 73.372; scale 0.296

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	158	100.0	21	38	W67579	T-cell epitope peptid	4.34e-09
2	158	100.0	21	2	R11896	Immunogenic conjugate	4.34e-09
3	158	100.0	21	36	W67034	Tetanus toxin fragmen	4.34e-09
4	158	100.0	21	20	W06130	Tetanus toxoid protei	4.34e-09
5	158	100.0	21	28	W46449	Broad range helper T	4.34e-09
6	158	100.0	21	17	R88397	T-cell antigen TR3 pe	4.34e-09
7	158	100.0	21	37	W73222	Tetanus toxoid epitop	4.34e-09
8	158	100.0	32	13	R62702	LHRH-containing immu	4.34e-09
9	158	100.0	63	3	R14263	Immunogenic branched	4.34e-09
10	158	100.0	64	3	R14261	Immunogenic branched	4.34e-09
11	158	100.0	65	3	R14265	Immunogenic branched	4.34e-09
12	158	100.0	65	3	R14262	Immunogenic branched	4.34e-09
13	158	100.0	77	3	R14264	Immunogenic branched	4.34e-09
14	158	100.0	452	2	R12471	Tetanus toxin fragmen	4.34e-09
15	158	100.0	573	2	P70345	Portion of B fragment	4.34e-09
16	158	100.0	618	33	W48909	SOD-1/TIC hybrid prot	4.34e-09

17	152	96.2	22	16	R82574	Tetanus toxin helper	2.09e-08
18	152	96.2	22	13	R62693	Helper T cell epitope	2.09e-08
19	152	96.2	22	19	W05600	Tetanus toxin helper	2.09e-08
20	152	96.2	32	16	R82597	IgE CH4 region contg.	2.09e-08
21	152	96.2	34	16	R83562	IgE CH4 region contg.	2.09e-08
22	135	85.4	20	35	W71322	Universal helper T-ce	1.71e-06
23	132	83.5	20	34	W69279	Tetanus toxin fragmen	3.68e-06
24	129	81.6	21	27	W35438	T-cell stimulatory pe	7.92e-06
25	116	73.4	14	35	W78832	Tetanus toxoid protei	2.12e-04
26	114	72.2	15	35	W78833	Tetanus toxoid protei	3.50e-04
27	114	72.2	22	20	W03002	Carrier peptide for a	3.50e-04
28	95	60.1	473	35	W68400	Clostridium botulinum	3.73e-02
29	93	58.9	144	20	W09015	Immunogenic type F bo	6.03e-02
30	93	58.9	431	20	W09014	Immunogenic type F bo	6.03e-02
31	93	58.9	448	35	W68399	Clostridium botulinum	6.03e-02
32	93	58.9	472	35	W68393	Clostridium botulinum	6.03e-02
33	93	58.9	472	35	W68394	Clostridium botulinum	6.03e-02
34	93	58.9	1169	30	W56017	Recombinant botulinum	6.03e-02
35	93	58.9	1291	35	W68392	Clostridium botulinum	6.03e-02
36	90	57.0	438	17	R95008	Type A neurotoxin C f	1.24e-01
37	90	57.0	438	35	W68389	Clostridium botulinum	1.24e-01
38	90	57.0	445	35	W68391	Clostridium botulinum	1.24e-01
39	90	57.0	462	17	R95009	Type A neurotoxin C f	1.24e-01
40	90	57.0	462	35	W68390	Clostridium botulinum	1.24e-01
41	90	57.0	1296	17	R95010	C. botulinum type A n	1.24e-01
42	88	55.7	451	35	W68395	Clostridium botulinum	1.99e-01
43	88	55.7	452	35	W68396	Clostridium botulinum	1.99e-01
44	84	53.2	260	12	R56546	Partial sequence of n	5.10e-01
45	84	53.2	464	2	R11834	Non-structural protei	5.10e-01

ALIGNMENTS

RESULT 1
ID W67579 standard; peptide; 21 AA.
AC W67579;
DI 02-MAR-1999 (first entry)
DE T-cell epitope peptide #5 for chimeric fimbria/T-cell epitope peptide.
KW Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope;
KW Immunogenic composition; immune response.
OS Synthetic.
PN US5843464-A.
PD 01-DEC-1998.
PF 02-JUN-1995; 460502.
PR 02-JUN-1995; US-460502.
PA (OHIS) UNIV OHIO STATE.
PI Bakaletz LO, Kaumaya PTP;
DR WPI; 99-044514/04.
PT Synthetic chimeric fimbria peptide - useful for vaccination against
PT non-typable Haemophilus influenzae
PS Disclosure; Column 4; 16pp; English.
CC The invention relates to the manufacture of a synthetic chimeric peptide
CC comprising a non-typable Haemophilus influenzae fimbria peptide fused via
CC a linker peptide to a T-cell epitope peptide. The chimeric peptide is
CC used in immunogenic compositions which induce an immune response against
CC non-typable Haemophilus influenzae. This sequence represents an example
CC of a T-cell epitope peptide used to generate the chimeric peptide.
SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 38; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfwlrpkvsashle 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 2
ID R11896 standard; peptide; 21 AA.
AC R11896;
DI 19-JUL-1991 (first entry)
DE Immunogenic conjugate constituent peptide, TR3.

KW Malaria vaccine; major histocompatibility complex.
OS Tetanus toxin.
FH Key Location/Qualifiers
FT peptide 1..14
FT /label= active fragment (claimed)
PN EP-427347-A.
PD 15-MAY-1991.
PF 07-NOV-1990; 202948.
PR 10-NOV-1989; IT-022355.
PA (ENTE) ENRICECERHE SPA.
PI Bianchi E, Pessi A, Corradin G;
DR WPI; 91-141874/20.
PT Synthetic peptide(s) used as universal carriers - for preparing
PT immunogenic conjugates used as vaccines against Plasmodium
PT falciparum
PS Claim 1; page 13; 16pp; English.
CC This peptide corresponds to residues 947-967 of Tetanus toxin. It can
CC be used as a universal carrier for the prepn. of an immunogenic
CC conjugate. It is covalently bound to a peptide or polysaccharide
CC hapten derived from a pathogen. This conjugate can be used as
CC a vaccine for malaria. This peptide is recognised by different T-
CC helper cell clones in association with alleles of the human MHC.
CC It contains 2 epitopes: (a) 953-967, recognised by DR5-restricted
CC clones; and (b) 947-960, recognised by all other DR and DP-
CC restricted clones.
SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfvlrvpkvsashle 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 3
ID W67034 standard; peptide; 21 AA.
AC W67034;
DT 15-DEC-1998 (first entry)
DE Tetanus toxin fragment (residues 947-967).
KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KW dimeric poly-lysine; epitope; tumour.
OS Clostridium tetani.
PN W09843677-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; E01922.
PR 27-MAR-1997; US-041726.
PA (INSP) INST PASTEUR.
PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;
DR WPI; 98-557071/47.
PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT with dimeric poly-lysine enabling multiple epitopes to be
PT covalently attached
PS Disclosure; Page 13; 55pp; English.
CC The invention relates to a new carbohydrate peptide conjugate, which
CC comprises a carrier with a dimeric poly-lysine enabling multiple
CC epitopes to be covalently attached to it. Also claimed are: (1) an
CC antibody purified from biological fluid or cells of organisms
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC kit comprising antigen-specific antibodies elicited by immunisation with
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC diagnosis kit are used to provide pharmaceutical compositions and
CC vaccines against tumours. These can be used to support an immune response
CC against viral infections caused by hepatitis virus, HIV or cytomegalo
CC virus. They can be used to enhance immune responses, especially B- and T-
CC cell responses, of humans and animals against bacterial infections. The
CC carbohydrate peptide conjugate stimulates the antibody and T-cell
CC response without stimulating undesired immune responses. The composition
CC is capable of increasing the survival of tumour bearing humans and
CC animals. The present sequence corresponds to residues 947-967 of tetanus
CC toxin. The synthetic peptide corresponding to this sequence may be used
as epitope in a carbohydrate peptide conjugate.

SQ Sequence 21 AA;
Query Match 100.0%; Score 158; DB 36; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfvlrvpkvsashle 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
ID W06130 standard; Peptide; 21 AA.
AC W06130;
DT 07-FEB-1997 (first entry)
DE Tetanus toxoid protein T-cell epitope.
KW Cholesteryl ester transfer protein; CERP; antigen; vaccine;
KW cardiovascular disease; atherosclerosis; tetanus toxoid;
KW T-cell epitope.
OS Clostridium tetani.
PN W09634888-A1.
PD 07-NOV-1996.
PF 01-MAY-1996; U06147.
PR 01-MAY-1995; US-432483.
PA (TCEL-) T CELL SCI INC.
PI Rittershaus CW, Thomas LJ;
DR WPI; 96-506103/50.
PT Cholesteryl ester transfer protein B cell epitope linked to T cell
PT epitope - used to generate vaccine to regulate CERP activity for
PT decreasing the risk of developing a cardiovascular disease e.g.
PT atherosclerosis
PS Claim 11; Page 43; 72pp; English.
CC A helper T-cell epitope (W06130) comprises amino acids 947-967
CC of tetanus toxoid protein. It can be utilised in novel peptide
CC vaccines (see also W06129, W06132) also including B-cell
CC epitope(s) from human or rabbit cholesteryl ester transfer
CC protein (CERP) to elicit an immune response against endogenous
CC CERP activity, thereby treating or preventing a cardiovascular
CC disease, such as atherosclerosis.
SQ Sequence 21 AA;

Query Match 100.0%; Score 158; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfvlrvpkvsashle 21
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 5
ID W46449 standard; Peptide; 21 AA.
AC W46449;
DT 18-MAY-1998 (first entry)
DE Broad range helper T cell epitope from the tetanus toxoid protein.
KW Cholesteryl ester transfer protein; CERP; cholesteryl ester;
KW high density lipoprotein; HDL; very low density lipoprotein; VLDL;
KW low density lipoprotein; LDL; T cell epitope; antibody;
KW DNA plasmid-based vaccine; broad range helper T cell epitope;
KW treatment; cardiovascular disease.
OS Clostridium tetani.
PN W09741227-A1.
PD 06-NOV-1997.
PF 01-MAY-1997; U07294.
PR 21-FEB-1997; US-802967.
PR 01-MAY-1996; US-640713.
PA (TCEL-) T CELL SCI INC.
PI Thomas LJ;
DR WPI; 97-549731/50.
PT DNA plasmid-based vaccine encodes CERP B cell and helper T cell
PT epitope(s) - used for elevating high density lipoprotein levels, and
PT for treating cardiovascular disease
PS Disclosure; Page 44; 67pp; English.

CC The present sequence represents a broad range helper T cell epitope
CC of the tetanus protein. It can be used in DNA plasmid-based vaccines
CC against cholesterol ester transfer proteins (CEPps). CEPps mediate the
CC transfer of cholesterol esters from high density lipoprotein (HDL)
CC to very low density lipoprotein (VLDL) and low density lipoprotein (LDL),
CC and vice versa. An increased CERP activity produces an atherogenic
CC lipoprotein profile and induces atherosclerosis. A DNA plasmid-based
CC vaccine comprises sequences encoding at least one B cell epitope of CERP
CC linked in frame with at least one segment encoding a broad range helper
CC T cell epitope. The vaccines can be used to elevate the ratio of
CC circulating HDL to circulating LDL, VLDL or total cholesterol in a human.
CC It can also be used for decreasing the level of endogenous CERP activity
CC in a human. The vaccine can be used to produce anti-CERP antibodies in
CC vivo and for treating cardiovascular disease.

CC Sequence 21 AA;

Query Match 100.0%; Score 158; DB 28; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvswlrpkyasashle 21
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 6
ID R88397 standard; Peptide; 21 AA.
AC R88397;
DI 12-JUN-1996 (first entry)
DE T-cell antigen IT3 peptide.
KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;
KW alpha-helix; coiled-coil heterodimer; core peptide; subunit.
OS Synthetic.
PN WO9531480-A1.
PD 23-NOV-1995.
PF 18-MAY-1995; CA0293.
PR 18-MAY-1994; US-245507.
PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
PI Cachia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;
PI Zhou NE;
DR WPI; 96-010880/01.
PT Hetero:dimeric polypeptide immunogen in coiled-coil configuration
PT with different antigens on each sub:unit - useful in vaccines and
PT for antibody prodn.
PS Claim 7; Page 62; 95pp; English.
CC This T-cell antigen IT3 peptide may be attached to a core peptide
CC contained in one of the 2 subunits of an alpha-helical coiled-coil
CC heterodimer. Each core peptide is comprised of terminal and
CC internal AA repeat sequences. This peptide antigen is attached
CC to the core peptide through covalent linkages to certain AA of the
CC internal repeats. The 2 subunits of the heterodimer are arranged
CC in a stable alpha-helical coiled-coil configuration having a 1:1
CC stoichiometry, and the peptide antigen is disposed toward the outer
CC surfaces of the configuration. The heterodimer may be used as a
CC synthetic vaccine (optionally multivalent) or to generate
CC antibodies.
CC Sequence 21 AA;

Query Match 100.0%; Score 158; DB 17; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvswlrpkyasashle 21
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 7
ID W73222 standard; Protein; 21 AA.
AC W73222;
DI 25-JAN-1999 (first entry)
DE Tetanus toxoid epitope.
KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;

KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
KW epidermal growth factor receptor; breast cancer; ovarian cancer.
OS Synthetic.
PN US5837243-A.
PD 17-NOV-1998.
PF 07-JUN-1996; 661052.
PR 07-JUN-1996; US-661052.
PR 07-JUN-1995; US-484172.
PA (MEDA-) MEDAREX INC.
PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
DR WPI; 99-023374/02.
PT Specific killing of tumour cells - using a multi-specific molecule
PT comprising an anti-Fc receptor antibody and a portion which binds to
PI a target cell
PS Example 7; Column 27; 57pp; English.
CC This sequence represents a tetanus toxoid epitope and is recognised
CC by the multispecific single chain antibody designated H22. The
CC antibody can be used in the method of the invention for inducing
CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
CC which is characterised by overexpression of HER 2/neu or epidermal growth
CC factor receptor (EGFR), comprises contacting the tumour cell with a
CC multispecific protein molecule (preferably a single chain antibody)
CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
CC binds to EGFR. The method can be used for treating cancers especially
CC breast cancer or ovarian cancer. The multispecific antibody can also
CC be administered prophylactically to vaccinate a subject against infection
CC by a target cell.
CC Sequence 21 AA;

Query Match 100.0%; Score 158; DB 37; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvswlrpkyasashle 21
QY 1 FNNFTVSEWLRVPKVSASHLE 21

RESULT 8
ID R62702 standard; peptide; 32 AA.
AC R62702;
DI 10-SEP-1995 (first entry)
DE LHRH-containing immunogenic peptide.
KW Helper T cell epitope; universal immune stimulator; invasin; hapten;
KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;
KW androgen-dependent carcinoma; antitumour; infertility;
KW tetanus toxin.
OS Synthetic.
FH Key Location/Qualifiers
FT domain 1..22
FT domain /note= "tetanus toxin helper T cell epitope"
FT domain 23..32
FT domain /note= "LHRH hapten"
PN WO9425060-A.
PD 10-NOV-1994.
PF 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
PI Ladd AE, Wang CY, Zamb T;
DR WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PT that suppress LHRH activity in males and females
PS Claim 8; Page 84; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein hapten containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC potent immune responses to the coupled peptide or protein. The
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
CC which elicits an immune response to the coupled peptide in members of

CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence from the invasive protein of Yersinia.
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
CC invasive and Th domains and between the immune stimulator and hapten
CC components. When the hapten is LHRH, then optionally the invasive domain
CC can be omitted from the immune stimulator component.
CC The present sequence represents an LHRH-containing, invasive-free
CC immunogenic peptide as above which can be used as a potent vaccine for
CC treating e.g. prostatic hyperplasia, androgen-dependent carcinoma,
CC prostatic carcinoma, testicular carcinoma, endometriosis, benign
CC uterine tumours, recurrent functional ovarian cysts, (severe)
CC premenstrual syndrome or oestrogen-dependent breast cancer, or for
CC induction of infertility.
SQ Sequence 32 AA;

Query Match 100.0%; Score 158; DB 13; Length 32;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 fnftvsfwlrpkvsashle 23
|||||
1 FNNFTVSFWLRPKVSASHLE 21

RESULT 9
ID R14263 standard; Peptide; 63 AA.
AC R14263;
DT 14-JAN-1992 (first entry)
DE Immunogenic branched polypeptides for antimalarial vaccines.
KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
OS Synthetic.
FH Key Location/Qualifiers
FT region 25..38
FT /label= T epitope
FT 39..59
FT /label= T epitope
FT 60
FT modified_site
FT /note= "epsilon amino substituted with the sequence
FT (NANP)6QYIKANSKFTIGITEFNNFTVSFWLRPKVSASHLE"
FT 61
FT /note= "epsilon amino substituted by Lys in which
FT both alpha and epsilon amino groups are
FT substituted with the sequence
FT (NANP)6QYIKANSKFTIGITEFNNFTVSFWLRPKVSASHLE"
FT 62
FT modified_site
FT /note= "epsilon amino substituted by Lys in which
FT EP-450715-A.
FT 09-OCT-1991.
FT 28-MAR-1991; 200727.
FT 02-APR-1990; IT-019914.
PA (ENIE) ENIRICERCHE SPA.
PI Pessi A, Bianchi E, Corradin G;
DR WPI; 91-297504/41.
PT New immunogenic branched polypeptide derivs. - used as antigens
PT in enzyme immunoassays and as anti sporozoite vaccines against
PT Plasmodium falciparum
PS Claim 10; Page 15; 22pp; English.
CC each of the alpha and epsilon amino groups
CC is substituted by Lys, both of the latter
CC two Lys residues being substituted in each
CC of their alpha and epsilon amino groups by
CC (NANP)6QYIKANSKFTIGITEFNNFTVSFWLRPKVSASHLE"
CC The peptide is a specific example of highly generic
CC immunogenic substituted lysines or polylysines having a number n
CC (where n is 1-15) of L-lysine amino acid residues of alpha and
CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
CC and/or (n+1)/2 of the epsilon amino groups are substituted with
CC polypeptides consisting of one or more plasmodial B epitopes
CC covalently bound to one or more peptides with an amino acid sequence
CC corresponding to that of a T epitope such as FNNFTVSFWLRPKVSASHLEA
CC or QYIKANSKFTIGITE
CC The branched polypeptides can be used as immunogens for
CC preparing genetically non-restricted antimalaria vaccines and for

CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
CC samples. Determination can be effected by ELISA.
CC See also R14261-2, R14264-5 and R15436.
SQ Sequence 63 AA;

Query Match 100.0%; Score 158; DB 3; Length 63;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 39 fnftvsfwlrpkvsashle 59
|||||
1 FNNFTVSFWLRPKVSASHLE 21

RESULT 10
ID R14261 standard; peptide; 64 AA.
AC R14261;
DT 14-JAN-1992 (first entry)
DE Immunogenic branched polypeptides for antimalarial vaccines.
KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1..21
FT modified_site
FT /label= T epitope
FT 62
FT /note= "epsilon-amino substituted with the sequence
FT FNNFTVSFWLRPKVSASHLE(NANP)10 "
FT 63
FT modified_site
FT /note= "epsilon-amino group substituted with Lys in
FT which both alpha and epsilon amino groups
FT are substituted with the sequence
FT FNNFTVSFWLRPKVSASHLE(NANP)10"
FT 64
FT modified_site
FT EP-450715-A.
FT 09-OCT-1991.
FT 28-MAR-1991; 200727.
FT 02-APR-1990; IT-019914.
PA (ENIE) ENIRICERCHE SPA.
PI Pessi A, Bianchi E, Corradin G;
DR WPI; 91-297504/41.
PT New immunogenic branched polypeptide derivs. - used as antigens
PT in enzyme immunoassays and as anti sporozoite vaccines against
PT Plasmodium falciparum
PS Claim 8; Page 15; 22pp; English.
CC The peptide is a specific example of highly generic
CC immunogenic substituted lysines or polylysines having a number n
CC (where n is 1-15) of L-lysine amino acid residues of alpha and
CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
CC and/or (n+1)/2 of the epsilon amino groups are substituted with
CC polypeptides consisting of one or more plasmodial B epitopes
CC covalently bound to one or more peptides with an amino acid sequence
CC corresponding to that of a T epitope such as FNNFTVSFWLRPKVSASHLEA
CC or QYIKANSKFTIGITE
CC The branched polypeptides can be used as immunogens for
CC preparing genetically non-restricted antimalaria vaccines and for
CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
CC samples. Determination can be effected by ELISA.
CC See also R14262 - R14265 and R15436.
SQ Sequence 64 AA;

Query Match 100.0%; Score 158; DB 3; Length 64;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfwlrpkvsashle 21
|||||
1 FNNFTVSFWLRPKVSASHLE 21

RESULT 11
ID R14265 standard; Peptide; 65 AA.
AC R14265;
DT 14-JAN-1992 (first entry)
DE Immunogenic branched polypeptides for antimalarial vaccines.


```
KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1..21
FT modified_site /label= T epitope
FT 63
FT /note= "epsilon amino substituted with the sequence
FT FNNFTVSFWLRVPKVSASHLE(NANP)10K "
FT 64
FT modified_site
FT /note= "epsilon amino substituted with Lys in which
FT alpha and epsilon amino groups are each
FT substituted with the sequence
FT FNNFTVSFWLRVPKVSASHLE(NANP)10K"
PN EP-450715-A.
PD 09-OCT-1991.
PF 28-MAR-1991; 200727.
PR 02-APR-1990; IT-019914.
PA (ENIE ) ENRICEACHE SPA.
PI Pessi A, Bianchi E, Corradin G;
DR WPI; 91-297504/41.
PT New immunogenic branched polypeptide derivs. - used as antigens
PT in enzyme immunoassays and as anti sporozoite vaccines against
PT Plasmodium falciparum
PS Claim 13; Page 16; 22pp; English.
CC The peptide is a specific example of highly generic
CC immunogenic substituted lysines or polylysines having a number n
CC (where n is 1-15) of L-lysine amino acid residues of alpha and
CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
CC and/or (n+1)/2 of the epsilon amino groups are substituted with
CC polypeptides consisting of one or more plasmodial B epitopes
CC covalently bound to one or more peptides with an amino acid sequence
CC corresponding to that of a T epitope such as FNNFTVSFWLRVPKVSASHLEA
CC or QYKANSKFIGITE .
CC The branched polypeptides can be used as immunogens for
CC preparing genetically non-restricted antimalaria vaccines and for
CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
CC samples. Determination can be effected by ELISA.
CC See also R14261 - R14264, R14266 and R15436.
SQ Sequence 65 AA;

Query Match 100.0%; Score 158; DB 3; Length 65;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 fnftvsfwlrvpkvsashle 21
   ||||||||||||||||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 12
ID R14262 standard; Peptide; 65 AA.
AC R14262;
DT 14-JAN-1992 (first entry)
DE Immunogenic branched polypeptides for antimalarial vaccines.
KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
OS Synthetic.
FH Key Location/Qualifiers
FT region 1..21
FT modified_site /label= T epitope
FT 62
FT /note= "epsilon amino substituted by the sequence
FT VOGEEESNDK"
FT 63
FT modified_site
FT /note= "epsilon amino substituted by Lys in which the
FT alpha amino is substituted with the sequence
FT FNNFTVSFWLRVPKVSASHLE(NANP)10 and the epsilon
FT amino is substituted with the sequence
FT VOGEEESNDK"
FT 64
FT modified_site
FT /note= "epsilon amino substituted by Lys in which
FT both the alpha and epsilon amino groups are
FT substituted with further Lys residues,the
FT latter two Lys residues each being substituted
```

```
FT on the alpha amino by FNNFTVSFWLRVPKVSASHL-
FT E(NANP)10 and on the epsilon amino by the
FT sequence VOGEEESNDK"
PN EP-450715-A.
PD 09-OCT-1991.
PF 28-MAR-1991; 200727.
PR 02-APR-1990; IT-019914.
PA (ENIE ) ENRICEACHE SPA.
PI Pessi A, Bianchi E, Corradin G;
DR WPI; 91-297504/41.
PT New immunogenic branched polypeptide derivs. - used as antigens
PT in enzyme immunoassays and as anti sporozoite vaccines against
PT Plasmodium falciparum
PS Claim 9; Page 15; 22pp; English.
CC The peptide is a specific example of highly generic
CC immunogenic substituted lysines or polylysines having a number n
CC (where n is 1-15) of L-lysine amino acid residues of alpha and
CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
CC and/or (n+1)/2 of the epsilon amino groups are substituted with
CC polypeptides consisting of one or more plasmodial B epitopes
CC covalently bound to one or more peptides with an amino acid sequence
CC corresponding to that of a T epitope such as FNNFTVSFWLRVPKVSASHLEA
CC or QYKANSKFIGITE.
CC The branched polypeptides can be used as immunogens for
CC preparing genetically non-restricted antimalaria vaccines and for
CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
CC samples. Determination can be effected by ELISA.
CC See also R14262 - R14265 and R15436.
SQ Sequence 65 AA;
```

```
Query Match 100.0%; Score 158; DB 3; Length 65;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 1 fnftvsfwlrvpkvsashle 21
   ||||||||||||||||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21
```

```
RESULT 13
ID R14264 standard; Peptide; 77 AA.
AC R14264;
DT 14-JAN-1992 (first entry)
DE Immunogenic branched polypeptides for antimalarial vaccines.
KW Immunogen; Plasmodium; malaria; lysine; immunoassay.
OS Synthetic.
FH Key Location/Qualifiers
FT region 13..33
FT modified_site /label= T epitope
FT 75
FT /note= "epsilon amino substituted with the sequence
FT (NANP)3FNNFTVSFWLRVPKVSASHLE(NANP)10K"
FT 76
FT modified_site
FT /note= "epsilon amino substituted with Lys in which
FT alpha and epsilon amino groups are each
FT substituted with the sequence
FT (NANP)3FNNFTVSFWLRVPKVSASHLE(NANP)10K"
PN EP-450715-A.
PD 09-OCT-1991.
PF 28-MAR-1991; 200727.
PR 02-APR-1990; IT-019914.
PA (ENIE ) ENRICEACHE SPA.
PI Pessi A, Bianchi E, Corradin G;
DR WPI; 91-297504/41.
PT New immunogenic branched polypeptide derivs. - used as antigens
PT in enzyme immunoassays and as anti sporozoite vaccines against
PT Plasmodium falciparum
PS Claim 11; Page 16; 22pp; English.
CC The peptide is a specific example of highly generic
CC immunogenic substituted lysines or polylysines having a number n
CC (where n is 1-15) of L-lysine amino acid residues of alpha and
CC epsilon amide linkage, where (n+1)/2 of the alpha amino groups
CC and/or (n+1)/2 of the epsilon amino groups are substituted with
```

CC polypeptides consisting of one or more plasmidial B epitopes
CC covalently bound to one or more peptides with an amino acid sequence
CC corresponding to that of a T epitope such as FNNFTVSFWLRVPKVSASHLEA
CC or QYIKANSKFIGITE .
CC The branched polypeptides can be used as immunogens for
CC preparing genetically non-restricted antimalaria vaccines and for
CC determining anti-Plasmodium antibodies in blood, serum and blood-spot
CC samples. Determination can be effected by ELISA.
CC See also R14261 - R14263, R14265 and R15436.
SQ Sequence 77 AA;

Query Match 100.0%; Score 158; DB 3; Length 77;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 13 fnftvsfwlrvpkvsashle 33
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 14
ID R12471 standard; Protein; 452 AA.
AC R12471;
DT 05-AUG-1991 (first entry)
DE Tetanus toxin fragment C encoded by gene with increased G+C content.
KW Terminator; vaccine.
OS Synthetic.
PN EP-430645-A.
PD 05-JUN-1991.
PF 27-NOV-1990; 312870.
PR 28-NOV-1989; GB-026832.
PR 17-MAR-1990; GB-006097.
PA (WELL) WELLCOME FOUNDATION LTD.
PI Makof A.J. Romanos MA, Clare JJ, Fairweather NF;
DR WPI: 91-165115/23.
DR N-PSDB; Q12121.
PT DNA sequence encoding tetanus toxin fragment C - useful in the
PT manufacture of vaccines for immunity to tetanus utilising yeast
PT as host organism.
PS Disclosure; Fig 2; 50pp; English.
CC The (G+C) content of the synthetic gene is increased by 47% wrt the
CC native sequence. This eliminates six "terminator" regions which
CC were found to be present in (A+T) rich regions. The terminators
CC (termination/endo-nucleolytic processing/polyadenylation sites)
CC were previously responsible for incomplete transcription of the
CC mRNA. The elimination of these elements (using codon degeneracy)
CC provided for successful expression in yeast of the tetanus toxin
CC fragment C.
SQ Sequence 452 AA;

Query Match 100.0%; Score 158; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 fnftvsfwlrvpkvsashle 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 15
ID P70345 standard; Protein; 573 AA.
AC P70345;
DT 22-APR-1991 (first entry)
DE Portion of B fragment and all of the C fragment of tetanus toxin.
KW TT; vaccine.
OS Clostridium tetani.
PN EP-209281-A.
PD 21-JAN-1987.
PF 27-JUN-1986; 305029.
PR 28-JUN-1985; GB-016442.
PA (WELL) Wellcome Foundation Ltd.
PI Fairweather NF;
DR WPI: 87-015999/03.

DR N-PSDB; N70545.
PT Cloned DNA sequence coding for tetanus toxin - or its fragments
PT contg. epitope used to express antigens for vaccine production.
PS Claim 4; Fig 1; 36pp; English.
CC Gene product comprises a tetanus toxin fragment, which may be
CC expressed in a transformed host, and used as an antigen in vaccine
CC production, against the disease.
SQ Sequence 573 AA;

Query Match 100.0%; Score 158; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 205 fnftvsfwlrvpkvsashle 225
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 16
ID W48909 standard; Protein; 618 AA.
AC W48909;
DT 23-SEP-1998 (first entry)
DE SOD-1/TTC hybrid protein.
KW Chimeric; copper-zinc superoxide dismutase; SOD-1; TTC; SOD:Tet451;
KW tetanus toxin fragment C; tetanus holotoxin; nerve cell; stroke;
KW neurological disorder; oxidative stress; brain hypoxia-reperfusion;
KW epilepsy; Parkinson's disease; Huntington's disease.
OS Chimeric - Homo sapiens.
OS Chimeric - Clostridium tetani.
FH Key
FT Region Location/Qualifiers
FT 1..163 /note= "SOD-1"
FT 168..618 /note= "TTC moiety"
FT Region

US5780024-A.
PD 14-JUL-1998.
PF 21-JUN-1996; 668381.
PR 23-JUN-1995; US-000473.
PR 21-JUN-1996; US-668381.
PA (GEHO) GEN HOSPITAL CORP.
PA (UYMA-) UNIV MARYLAND BALTIMORE.
PI Brown RH, Fishman PS, Francis JW, Hosler BA;
DR WPI: 98-412999/35.
DR N-PSDB; V32580.
PT New hybrid protein of superoxide dismutase and tetanus toxin
PT fragment C - having increased uptake by neurons and retention of
PT enzymatic activity in these cells, for treating neurological
PT diseases associated with oxidative stress
PS Claim 7; Columns 23-26; 23pp; English.
CC The present sequence represents an enzymatically active human
CC copper-zinc superoxide dismutase (SOD-1) fused at its carboxyl
CC terminus with the tetanus toxin fragment C (TTC) moiety. The TTC
CC moiety constitutes amino acid residues 865-1315 of the tetanus
CC holotoxin. The hybrid protein, referred as SOD:Tet451, is claimed
CC to have the following properties: (a) it exhibits Cu/Zn SOD enzymatic
CC activity; (b) the TTC moiety selectively binds to nerve cells and
CC allows uptake of the hybrid protein into these cells; and (c) it
CC retains substantial SOD enzymatic activity following cellular uptake.
CC SOD:Tet451 is claimed to be useful for treating neurological disorders
CC associated with oxidative stress, e.g. stroke, brain hypoxia-reperfusion,
CC epilepsy, Parkinson's and Huntington's diseases.
SQ Sequence 618 AA;

Query Match 100.0%; Score 158; DB 3; Length 618;
Best Local Similarity 100.0%; Pred. No. 4.34e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 250 fnftvsfwlrvpkvsashle 270
QY 1 FNNFTVSFWLRVPKVSASHLE 21

Search completed: Wed Aug 4 15:48:30 1999

Thu Aug 5 09:01:22 1999

US-09-049-847-2.rag

Page 7

Job time : 18 secs.

This Page Blank (uspto)

Query Match	60.6%;	Score 63;	DB 2;	Length 899;
Best Local	Similarity	50.0%;	Pred. No. 5.01e-01;	
Matches	7;	Conservative	4;	Mismatches 3; Indels 0; Gaps 0;

```

Db      124 QYISNATFTGLSE 137
      ||| :| :| :|
QY      1 QYIKANSKFITGITE 14

```

RESULT	4
ENTRY	S77138
TITLE	hypothetical protein slr1882 - <i>Synechocystis</i> sp. (strain PCC

ORGANISM	#formal_name	Synechocystis sp.
#variety	PCC 6803	
DATE	25-Apr-1997	#sequence_revision 25-Apr-1997 #text_change 21-Aug-1998

ACCESSIONS	S77138
REFERENCE	S74322
#authors	Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;

#Journal
#title
DNA Res. (1996) 3:109-136
Sequence analysis of the genome of the unicellular
cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.

```
#cross-references MUID:97061201
#accession      S77138
```

```
##status      nucleic acid sequence not shown; translation not shown
##molecule_type  DNA
##residues      1-284  ##label KAN
##cross-references  EMBL:D90908; GB:AB001339; NID:g1652725; PID:d1018429
                    PID:g1652777
##note          the nucleotide sequence was submitted to the EMBL Data
                    Library, June 1996
```

```

GENETICS      GTG
#start_codon #superfamily conserved hypothetical protein H10963
CLASSIFICATION #length 284 #molecular-weight 31380 #checksum 39655
SUMMARY

```

Query Match	59.6%;	Score 62;	DB 2;	Length 284;
Best Local Similarity	33.3%;	Pred. No. 7.75e-01;		
Matches	5;	Conservative	8;	Mismatches 2;
				Indels 0;
				Gaps 0;

2b 248 HYLREPTEFAGLDOL 262
:|:::| | : :|
QY 1 QYIKANSKFIGITEL 15

RESULT	5	F69792	#type complete			
ENTRY		hypothetical protein yeeb	-	Bacillus subtilis		
TITLE		#formal_name Bacillus subtilis				
ORGANISM		05-Dec-1997	#sequence_revision	05-Dec-1997	#text_change	
DATE		24-Sep-1998				

ACCESSIONS
REFERENCE
#authors

unst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Graser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guisepi, G.; Guy, B.T.; Haga, K.; Haiech, J.; Harwood,

C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandendol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weltzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamamoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

Nature (1997) 390:249-256

The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

```

#status      preliminary; nucleic acid sequence not shown;
              translation not shown

```

```
##molecule_type DNA
##residues 1-598 ##label KUN
##cross-references GB:Z99107; GB:AL009126; NID:g2632866; PID:e1182657;
```

PID:g2632991
##experimental_source strain 168

GENETICS	yeab	#length	598	#molecular-weight	67963	#checksum	7652
#gene							
SUMMARY							

Query Match	59.6%;	Score 62;	DB 2;	Length 598;
Best Local	Similarity 40.0%;	Pred. No. 7.75e-01;		
Matches	6;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;

Db 531 EFIRMA DKFINIEDL 545
::: ||| | : |
QY 1 QYIKANSKEIGITEL 15

RESULT	6	
ENTRY	G64026	#type complete
TITLE	[acyl-carrier-protein] phosphodiesterase (EC 3.1.4.14) H11366	
	- Haemophilus influenzae	
ALTERNATE_NAMES	conserved hypothetical protein H11366	
ORGANISM	#formal_name Haemophilus influenzae	
DATE	18-Aug-1995	#sequence-revision 18-Aug-1995 #text_change
	21-Nov-1998	

ACCESSIONS
REFERENCE
#authors

#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.; FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton, M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, T.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.

#journal Science (1995) 269:496-512

#title Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.

#cross-references G64026

#accession G64026

```
##status      nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-194 ##label TIGR
##cross-references GB:U32816; GB:U42023; NID:g1574193; PID:g1574198;
               TIGR:HI1366
##experimental_source strain Rd KW20
FUNCTION
#description  catalyzes hydrolysis of the phosphopantetheine residue from
               holo-acyl-carrier-protein
CLASSIFICATION #superfamily acyl carrier protein phosphodiesterase
KEYWORDS       phosphoric diester hydrolase
SUMMARY        #length 194 #molecular-weight 21208 #checksum 8859

Query Match
Best Local Similarity 58.7%; Score 61; DB 2; Length 194;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Db 147 QYKXSIIGFIGITDV 161
QY 1 QYIKANSKFIGITEL 15

RESULT 7
ENTRY 7
TITLE  S05532 #type complete
ALTERNATE_NAMES gamma-glutamyltransferase (EC 2.3.2.2) - pig
ORGANISM gamma-glutamyl transferidase
#formal_name Sus scrofa domestica #common_name domestic pig
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
      29-Jan-1999
ACCESSIONS S05532
REFERENCE S05532
#authors  Papandrikopoulou, A.; Frey, A.; Gassen, H.G.
#journal  Eur. J. Biochem. (1989) 183:693-698
#title    Cloning and expression of gamma-glutamyl transferidase from
           isolated porcine brain capillaries.
#cross-references MUID:89377838
#accession S05532
#status    not compared with conceptual translation
##molecule_type DNA
##residues 1-568 ##label PAP
CLASSIFICATION ##cross-references GB:Z46922; NID:g600818; PID:g600819
KEYWORDS       #superfamily gamma-glutamyltransferase
               aminoacyltransferase; glycoprotein; heterodimer; membrane
               protein
FEATURE
1-379
380-568
SUMMARY        #product gamma-glutamyltransferase heavy chain #status
               predicted #label HCH\
               #product gamma-glutamyltransferase light chain #status
               predicted #label LCH
               #length 568 #molecular-weight 61315 #checksum 3041

Query Match
Best Local Similarity 57.7%; Score 60; DB 2; Length 568;
Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 531 HYIQDASTFIGVQ 544
QY 1 QYIKANSKFIGITE 14

RESULT 8
ENTRY 8
TITLE  S31029 #type complete
ORGANISM gene 84 protein - Mycobacterium phage L5
DATE 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change
      09-Sep-1997
ACCESSIONS S31029
REFERENCE S30949
#authors  Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
#journal  Mol. Microbiol. (1993) 7:407-417
#title    Superinfection immunity of mycobacteriophage L5: applications
           for genetic transformation of mycobacteria.
#accession S31029
```

```
##status      nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-66 ##label DON
##cross-references EMBL:Z18946; NID:g15859; PID:e59702; PID:g579152
##note        the nucleotide sequence was submitted to the EMBL Data
               Library, December 1992
GENETICS
#gene 84
#start_codon GTG
SUMMARY        #length 66 #molecular-weight 7424 #checksum 8203

Query Match
Best Local Similarity 56.7%; Score 59; DB 2; Length 66;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 50 YIKRNGKRVG 59
QY 2 YIKANSKFIG 11

RESULT 9
ENTRY 9
TITLE  G48677 #type fragment
ALTERNATE_NAMES Ig heavy chain V-D-J region (419.1) - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 19-May-1994 #sequence_revision 19-May-1994 #text_change
      17-Mar-1999
ACCESSIONS G48677
REFERENCE G48677
#authors  Tassignon, J.; Brait, M.; Jamila, I.; Urbain, J.; Gottlieb,
           P.; Brown, A.; Hasemann, C.A.; Capra, J.D.; Meek, K.
           Proc. Natl. Acad. Sci. U.S.A. (1993) 90:9508-9512
           Molecular characterization of monoclonal CRT-A-positive
           anti-arsenate antibodies derived from idiotypic-negative
           mice bearing a light chain polymorphism.
#cross-references MUID:94022404
#accession G48677
#status    Preliminary; not compared with conceptual translation
##molecule_type mRNA
##residues 1-123 ##label TAS
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS       heterotetramer; immunoglobulin
FEATURE
15-98
SUMMARY        #domain immunoglobulin homology #label IVM
               #length 123 #checksum 1208

Query Match
Best Local Similarity 56.7%; Score 59; DB 2; Length 123;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Db 56 DYIKYNEKFKGTTL 70
QY 1 QYIKANSKFIGITEL 15

RESULT 10
ENTRY 10
TITLE  F70361 #type complete
ORGANISM tRNA pseudouridine 55 synthase - Aquifex aeolicus
DATE 08-May-1998 #sequence_revision 08-May-1998 #text_change
      21-Aug-1998
ACCESSIONS F70361
REFERENCE A70300
#authors  Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
           Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
           Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
           J.M.; Olson, G.J.; Swanson, R.V.
           Nature (1998) 392:353-358
           The complete genome of the hyperthermophilic bacterium
           Aquifex aeolicus.
#cross-references MUID:98196666
#accession F70361
#status    preliminary; nucleic acid sequence not shown;
           translation not shown
```

```
##molecule_type DNA
##residues 1-287 ##label AQF
##cross-references GB:AE000703; NID:g2983287; PID:g2983293; GB:AE000657
##experimental_source strain VF5
```

GENETICS

```
#gene trueB
CLASSIFICATION
SUMMARY
#length 287 #molecular-weight 32259 #checksum 7335
```

```
Query Match 56.7%; Score 59; DB 2; Length 287;
Best Local Similarity 80.0%; Pred. No. 2.79e+00;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Db 266 DSKFIGIGEL 275
QY 6 NSKFIGITEL 15
```

```
RESULT 11
ENTRY B71808 #type complete
TITLE type II restriction enzyme - Helicobacter pylori (strain J99)
ORGANISM #formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 12-Feb-1999
```

```
ACCESSIONS B71808
REFERENCE A71800
#authors Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
```

```
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
```

```
#cross-references MUID:99120557
#accession B71808
##status preliminary
##molecule_type DNA
##residues 1-326 ##label ARN
##cross-references GB:AE001566; GB:AE001439; NID:g4156051; PID:g4156064
##experimental_source strain J99
```

GENETICS

```
#gene jhp1442
SUMMARY
#length 326 #molecular-weight 39034 #checksum 2234
```

```
Query Match 55.8%; Score 58; DB 2; Length 326;
Best Local Similarity 50.0%; Pred. No. 4.24e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Db 43 YIQSNIKIYISLIPL 56
QY 2 YIKANSKFIGITEL 15
```

```
RESULT 12
ENTRY C69908 #type complete
TITLE delta-endotoxin homolog yokG - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998
```

```
ACCESSIONS C69908
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
```

M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, C.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott, G.; A.M.; Prescan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzeneegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

```
#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
```

```
#cross-references MUID:98044033
```

```
#accession C69908
```

```
##status
```

```
preliminary; nucleic acid sequence not shown;
translation not shown
```

```
##molecule_type DNA
```

```
##residues 1-357 ##label KUN
```

```
##cross-references GB:Z99115; GB:AL009126; NID:g2634478; PID:e1183607;
PID:g2634580
```

```
##experimental_source strain 168
```

GENETICS

```
#gene yokG
```

```
SUMMARY
#length 357 #molecular-weight 40742 #checksum 1804
```

```
Query Match 55.8%; Score 58; DB 2; Length 357;
Best Local Similarity 35.7%; Pred. No. 4.24e+00;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
```

```
Db 165 FLQGGNNFIGVTQL 178
```

```
QY 2 YIKANSKFIGITEL 15
```

```
RESULT 13
```

```
ENTRY F64690 #type complete
```

```
TITLE type IIS restriction enzyme R protein - Helicobacter pylori
(strain 26695)
```

```
ORGANISM #formal_name Helicobacter pylori
```

```
DATE 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 10-Oct-1997
```

```
ACCESSIONS F64690
```

```
REFERENCE A64520
```

```
#authors
```

Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.; Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzgerald, L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.C.; Utterback, I.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.M.; Venter, J.C.

```
Nature (1997) 388:539-547
```

```
The complete genome sequence of the gastric pathogen
```

```
#journal
```

```
#title
```

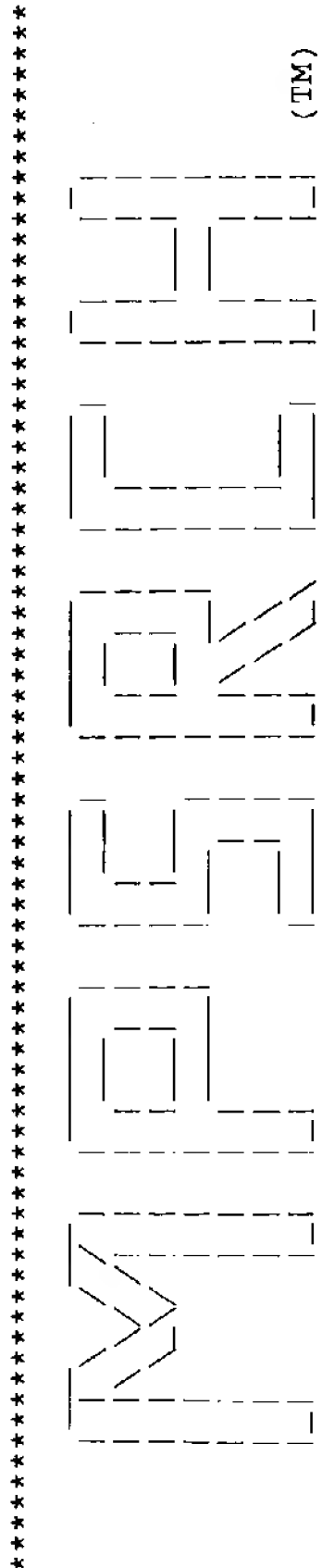


```

#cross-references MUID:97394467
#accession F64690
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-423 ##label TOM
##cross-references GB:AE000637; GB:AE000511; NID:g2314536; PID:g2314537;
TIGR:HP1366
SUMMARY #length 423 #molecular-weight 50047 #checksum 4262
Query Match 55.8%; Score 58; DB 2; Length 423;
Best Local Similarity 50.0%; Pred. No. 4.24e+00;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
Db 140 YIQSNIKYISLIPL 153
||:| |:|:|
QY 2 YIKANSKFIGITEL 15
RESULT 14
ENTRY #type complete
TITLE hypothetical protein yaaC homolog Vxpspt7_orf269 - Mycoplasma
pneumoniae (ATCC 29342) (SGC3)
ALTERNATE_NAMES hypothetical protein Vxpspt7_orf269
ORGANISM #formal_name Mycoplasma pneumoniae
#variety ATCC 29342
DATE 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S73999
REFERENCE S73327
#authors Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li,
B.C.; Herrmann, R.
#journal Nucleic Acids Res. (1996) 24:4420-4449
#title Complete sequence analysis of the genome of the bacterium
Mycoplasma pneumoniae.
#cross-references MUID:97105885
#accession S73999
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-269 ##label HIM
##cross-references EMBL:AE000062; GB:U00089; NID:g1674373; PID:g1674379
#note the nucleotide sequence was submitted to the EMBL Data
Library, November 1996
GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily conserved hypothetical protein HI0963
SUMMARY #length 269 #molecular-weight 30435 #checksum 2148
Query Match 54.8%; Score 57; DB 2; Length 269;
Best Local Similarity 40.0%; Pred. No. 6.40e+00;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
Db 235 QFIRPQQKFSGVQAL 249
|:|:| |:|:|
QY 1 QYIKANSKFIGITEL 15
RESULT 15
ENTRY #type complete
TITLE formylmethanofuran dehydrogenase (tungsten) (EC 1.2.99.-)
subunit C related protein - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS B64382
REFERENCE A64300
#authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; Fitzgerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
```

```

Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
Science (1996) 273:1058-1073
Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii.
#cross-references MUID:96337999
#accession B64382
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-322 ##label BUL
##cross-references GB:U67513; GB:L77117; NID:g1591365; PID:g1591371;
TIGR:MJ0658; PID:g1510742
GENETICS
#map_position FOR584512-585480
#start_codon TTG
KEYWORDS oxidoreductase
SUMMARY #length 322 #molecular-weight 36149 #checksum 2742
Query Match 54.8%; Score 57; DB 2; Length 322;
Best Local Similarity 61.5%; Pred. No. 6.40e+00;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
Db 153 IKGYRKFI SITEF 165
||:| |:|:|
QY 3 IKANSKFIGITEL 15
Search completed: Wed Aug 4 15:22:44 1999
Job time : 8 secs.
```

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:23:01 1999; MasPar time 3.13 Seconds
Tabular output not generated. 135.279 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 27.171; Variance 32.173; scale 0.845

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	104	100.0	1314	1	TETX_CLOTE	TETANUS TOXIN PRECURSOR	3.53e-12
2	63	60.5	213	1	KAD_MYCCA	ADENYLATE KINASE (EC 2	8.08e-02
3	63	60.5	899	1	V120_HSVSA	CAPSID ASSEMBLY PROTEI	8.08e-02
4	62	59.5	284	1	RIBF_SYNY3	RIBOFLAVIN KINASE (EC	1.34e-01
5	61	58.7	194	1	ACPD_HAEIN	ACYL CARRIER PROTEIN P	2.20e-01
6	60	57.7	568	1	GGT_PIG	GAMMA-GLUTAMYLTRANSPEP	3.61e-01
7	59	56.7	66	1	VG84_BPML5	GENE 84 PROTEIN (GP84)	5.87e-01
8	59	56.7	287	1	TRUB_AQUAE	TRNA PSEUDOURIDINE SYN	5.87e-01
9	57	54.8	269	1	RIBF_MYCPN	PUTATIVE RIBOFLAVIN KI	1.53e+00
10	57	54.8	333	1	DPOB_XENLA	DNA POLYMERASE BETA (E	1.53e+00
11	57	54.8	1630	1	MSPL_PLAFK	MEROZOITE SURFACE PROT	1.53e+00
12	57	54.8	1639	1	MSPL_PLAFW	MEROZOITE SURFACE PROT	1.53e+00
13	56	53.8	316	1	RIBC_BACSU	RIBOFLAVIN KINASE (EC	2.44e+00
14	56	53.8	601	1	PEPF_LACLA	OLIGOENDOPEPTIDASE F (2.44e+00
15	56	53.8	887	1	ACOC_CAEEL	PROBABLE ACONITATE HYD	2.44e+00
16	55	52.9	85	1	Y668_METJA	HYPOTHETICAL PROTEIN M	3.88e+00
17	55	52.9	461	1	NIFN_RHOCA	NITROGENASE IRON-MOLYB	3.88e+00
18	55	52.9	757	1	RTSL_YEAST	RTS1 PROTEIN (SCS1 PRO	3.88e+00
19	55	52.9	1182	1	ABL2_HUMAN	TYROSINE-PROTEIN KINAS	3.88e+00
20	54	51.9	501	1	DLDH_PEA	DIHYDROLIPOAMIDE DEHYD	6.13e+00
21	54	51.9	528	1	MDLC_PSEPU	BENZOYLFORMATE DECARBO	6.13e+00
22	54	51.9	644	1	YHJ9_YEAST	HYPOTHETICAL ALDEHYDE-	6.13e+00
23	53	51.0	97	1	VMT2_IAXI1	MATRIX PROTEIN M2.	9.61e+00

24	53	51.0	130	1	RS8_METJA	30S RIBOSOMAL PROTEIN	9.61e+00
25	53	51.0	176	1	NU6C_ARATH	NADH-PLASTOQUINONE OXI	9.61e+00
26	53	51.0	451	1	MJRD_BACSU	UDP-N-ACETYLMURAMOYLAL	9.61e+00
27	53	51.0	548	1	YDD3_SCHPO	HYPOTHETICAL 63.3 KD P	9.61e+00
28	53	51.0	1083	1	UL52_VZVD	HELICASE/PRIMASE COMPL	9.61e+00
29	53	51.0	1140	1	XPE_CERAE	POSSIBLE DNA-REPAIR PR	9.61e+00
30	53	51.0	1682	1	MSPI_PLAF3	MEROZOITE SURFACE PROT	9.61e+00
31	53	51.0	1701	1	MSPI_PLAFM	MEROZOITE SURFACE PROT	9.61e+00
32	53	51.0	1701	1	MSPI_PLAFF	MEROZOITE SURFACE PROT	9.61e+00
33	53	51.0	1726	1	MSPI_PLAFP	MEROZOITE SURFACE PROT	9.61e+00
34	53	51.0	1726	1	MSPI_PLAFC	MEROZOITE SURFACE PROT	9.61e+00
35	52	50.0	109	1	RPOI_METJA	DNA-DIRECTED RNA POLYM	1.50e+01
36	52	50.0	272	1	AAKC_HUMAN	5'-AMP-ACTIVATED PROTE	1.50e+01
37	52	50.0	447	1	GNTI_MOUSE	ALPHA-1,3-MANNOSYL-GLY	1.50e+01
38	52	50.0	528	1	V58K_BSMV	58 KD PROTEIN (BETA-B	1.50e+01
39	52	50.0	635	1	YGC3_YEAST	HYPOTHETICAL 70.6 KD P	1.50e+01
40	52	50.0	746	1	ABL_MLVAB	TYROSINE-PROTEIN KINAS	1.50e+01
41	52	50.0	774	1	RRP3_INCBE	RNA-DIRECTED RNA POLYM	1.50e+01
42	52	50.0	774	1	RRP3_INCUJ	RNA-DIRECTED RNA POLYM	1.50e+01
43	52	50.0	869	1	CFAC_ECOLI	CFA/I FIMBRIAL SUBUNIT	1.50e+01
44	52	50.0	1123	1	ABL_MOUSE	PROTO-ONCOGENE TYROSIN	1.50e+01
45	52	50.0	1130	1	ABL_HUMAN	PROTO-ONCOGENE TYROSIN	1.50e+01

ALIGNMENTS

RESULT	1	TETX_CLOTE	STANDARD;	PRT;	1314 AA.
ID	TETX_CLOTE				
AC	P04958;				
DT	13-AUG-1987 (REL. 05, CREATED)				
DT	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)				
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)				
DE	TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXILYSIN).				
OS	CLOSTRIDIUM TETANI.				
OC	PLASMID.				
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;				
OC	CLOSTRIDIUM.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE; 87053814.				
RA	EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,				
RA	WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;				
RT	"Tetanus toxin: primary structure, expression in E. coli, and				
RT	homology with botulinum toxins.";				
RL	EMBO J. 5:2495-2502(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CN3911;				
RX	MEDLINE; 87040747.				
RA	FAIRWEATHER N.F., LYNNESS V.A.;				
RT	"The complete nucleotide sequence of tetanus toxin.";				
RL	NUCLEIC ACIDS RES. 14:7809-7812(1986).				
RN	[3]				
RP	SEQUENCE OF 742-1314 FROM N.A.				
RX	MEDLINE; 86085672.				
RA	FAIRWEATHER N.F., LYNNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;				
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin				
RT	fragment C in Escherichia coli.";				
RL	J. BACTERIOL. 165:21-27(1986).				
RN	[4]				
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.				
RX	MEDLINE; 90201034.				
RA	KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;				
RT	"Arrangement of disulfide bridges and positions of sulphhydryl groups				
RT	in tetanus toxin.";				
RL	EUR. J. BIOCHEM. 188:39-45(1990).				
RN	[5]				
RP	PARTIAL SEQUENCE.				
RX	MEDLINE; 92037649.				
RA	KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;				
RT	"Limited proteolysis of tetanus toxin. Relation to activity and				
RT	identification of cleavage sites.";				
RL	EUR. J. BIOCHEM. 202:41-51(1991).				

RT "Primary structure of the herpesvirus saimiri genome.";
 RL J. VIROL. 66:5047-5058(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92230228.
 RA NICHOLAS J., CAMERON K.R., COLEMAN H., NEWMAN C., HONESS R.W.;
 RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
 RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
 RT organization between HVS and Epstein-Barr virus.";
 RL VIROLOGY 188:296-310(1992).
 CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL37,
 CC EBV-1 23, EBV BOLF1, VZV 21, HVS-1 63, AND HCMV UL47.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64346; G60384; -.
 DR EMBL; M86409; G330992; -.
 DR PIR; G36812; G36812.
 KW CAPSID ASSEMBLY.
 SQ SEQUENCE 899 AA; 103350 MW; C2D70154 CRC32;
 Query Match 60.6%; Score 63; DB 1; Length 899;
 Best Local Similarity 50.0%; Pred. No. 8.08e-02;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Db 124 QYITSNATFTGLSE 137
 ||| :|: | | :|
 QY 1 QYIKANSKFIGITE 14
 RESULT 4
 ID RIBF_SYNY3 STANDARD; PRT; 284 AA.
 AC P73651;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DI 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
 DE ADENYLYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
 DE SYNTHETASE).
 GN RIBF OR SLR1882.
 OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
 OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97061201.
 RA KANEKO I., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
 RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
 RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K.,
 RA OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A.,
 RA YAMADA M., YASUDA M., TABATA S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA RES. 3:109-136(1996).
 CC -!- CATALYTIC ACTIVITY: ATP + RIBOFLAVIN = ADP + FMN.
 CC -!- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D90908; G1652777; -.
 DR TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.
 KW

SQ SEQUENCE 284 AA; 31380 MW; A4BFCA0C CRC32;
 Query Match 59.6%; Score 62; DB 1; Length 284;
 Best Local Similarity 33.3%; Pred. No. 1.34e-01;
 Matches 5; Conservative 8; Mismatches 2; Indels 0; Gaps 0;
 Db 248 HYLRPETKFAGLDQL 262
 :|:|:|:| | :|
 QY 1 QYIKANSKFIGITEL 15
 RESULT 5
 ID ACPD_HAEIN STANDARD; PRT; 194 AA.
 AC P43013;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE ACYL CARRIER PROTEIN PHOSPHODIESTERASE (ACP PHOSPHODIESTERASE).
 GN ACPD OR H1366.
 OS HAEMOPHILUS INFLUENZAE.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PASTEURELLACEAE;
 OC HAEMOPHILUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD;
 RA CHANDLER M.S., SMITH R.A.;
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20;
 RX MEDLINE; 95350630.
 RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
 RA KERLAVAGE A.R., BULT C.J., IOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
 RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
 RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
 RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLUM E., COTTON M.D.,
 RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
 RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
 RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
 RA VENTER J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT influenzae Rd.";
 RT SCIENCE 269:496-512(1995).
 CC -!- FUNCTION: CONVERTS HOLO-ACP TO APO-ACP BY HYDROLYTIC CLEAVAGE OF
 CC THE PHOSPHOPANTETHEINE RESIDUE FROM ACP. ITS PHYSIOLOGICAL
 CC FUNCTION IS NOT CLEAR (BY SIMILARITY).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U20964; G687791; -.
 DR EMBL; U32816; G1574198; -.
 DR TIGR; H1366; -.
 KW HYDROLASE.
 SQ SEQUENCE 194 AA; 21208 MW; B4D866EF CRC32;
 Query Match 58.7%; Score 61; DB 1; Length 194;
 Best Local Similarity 53.3%; Pred. No. 2.20e-01;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
 Db 147 QYMKSIILGFIGITDV 161
 ||:|: | | | :|
 QY 1 QYIKANSKFIGITEL 15
 RESULT 6
 ID GGT_PIG STANDARD; PRT; 568 AA.
 AC P20735;


```

Best Local Similarity 80.0%; Pred. No. 5.87e-01;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 266 DSKFIGIGEL 275
QY 6 NSKFIGITEL 15

RESULT 9
ID RIBE_MYCPN STANDARD; PRT; 269 AA.
AC P75587;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PUTATIVE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
DE ADENYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
DE SYNTHETASE).
GN RIBF.
OS MYCOPLASMA PNEUMONIAE.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; MOLLICUTES;
OC MYCOPLASMATACEAE; MYCOPLASMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE; 97105885.
RA HIMMELREICH R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,
RA HERRMANN R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL NUCLEIC ACIDS RES. 24:4420-4449(1996).
CC -!- CATALYTIC ACTIVITY: ATP + RIBOFLAVIN = ADP + FMN.
CC -!- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000062; G1674379; -
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.
SQ SEQUENCE 269 AA; 30435 MW; 07AF8D7C CRC32;

Query Match 54.8%; Score 57; DB 1; Length 269;
Best Local Similarity 40.0%; Pred. No. 1.53e+00;
Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Db 235 QFIRPQKFSGVQAL 249
QY 1 QYIKANSKFIGITEL 15

RESULT 10
ID DPOB_XENLA STANDARD; PRT; 333 AA.
AC O57383;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE DNA POLYMERASE BETA (EC 2.7.7.7).
GN POLB.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98151235.
RA REICHENBERGER S., PFEIFFER P.;
RT "Cloning, purification and characterization of DNA polymerase beta
RT from Xenopus laevis -- studies on its potential role in DNA-end
RT joining.";
RL EUR. J. BIOCHEM. 251:81-90(1998).

```

```

-!- FUNCTION: REPAIR POLYMERASE. CONDUCTS "GAP-FILLING" DNA SYNTHESIS
CC IN A STEPWISE DISTRIBUTIVE FASHION RATHER THAN IN A PROGRESSIVE
CC FASHION AS FOR OTHER DNA POLYMERASES.
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBUNIT: MONOMER.
CC -!- IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES: ALPHA, BETA, GAMMA,
CC DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR DIFFERENT REACTIONS
CC OF DNA SYNTHESIS.
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y15732; E1198727; -
KW PROSITE; PS00522; DNA_POLYMERASE_X; 1.
KW TRANSFERASE; DNA-DIRECTED DNA POLYMERASE; DNA REPLICATION.
FT INIT_MET 0 0 BY SIMILARITY.
FT ACT_SITE 182 182 INVOLVED IN PRIMER BINDING
FT ACT_SITE 189 189 INVOLVED IN PRIMER BINDING
FT ACT_SITE 189 189 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
FT ACT_SITE 191 191 INVOLVED IN PRIMER BINDING
SQ SEQUENCE 333 AA; 38162 MW; 6948CECC CRC32;

Query Match 54.8%; Score 57; DB 1; Length 333;
Best Local Similarity 38.5%; Pred. No. 1.53e+00;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

Db 228 VKGDTKFMGVQQL 240
QY 3 IKANSKFIGITEL 15

RESULT 11
ID MSP1_PLAFK STANDARD; PRT; 1630 AA.
AC P04932;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
DE (PMSA) (P190).
GN MSP-1.
OS PLASMODIUM FALCIPARUM (ISOLATE K1 / THAILAND).
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86136024.
RA MACKAY M., GOMAN M., BONE N., HYDE J.E., SCAIFE J., CERIA U.,
RA STUNNENBERG H., BUJARD H.;
RT "Polymorphism of the precursor for the major surface antigens of
RT Plasmodium falciparum merozoites: studies at the genetic level.";
RL EMBO J. 4:3823-3829(1985).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RA PAN W., TOLLE R., BUJARD H.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC (POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X03371; G929798; -.
DR PIR; A25120; SAZQK1.
DR PFAM; PF00008; EGF; 1.
KW MALARIA; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
KW TRANSMEMBRANE; GPI-ANCHOR.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67 84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614 1630 MEMBRANE ANCHOR.
FT CARBOHYD 97 97 POTENTIAL.
FT CARBOHYD 259 259 POTENTIAL.
FT CARBOHYD 755 755 POTENTIAL.
FT CARBOHYD 759 759 POTENTIAL.
FT CARBOHYD 774 774 POTENTIAL.
FT CARBOHYD 835 835 POTENTIAL.
FT CARBOHYD 911 911 POTENTIAL.
FT CARBOHYD 955 955 POTENTIAL.
FT CARBOHYD 1049 1049 POTENTIAL.
FT CARBOHYD 1156 1156 POTENTIAL.
FT CARBOHYD 1165 1165 POTENTIAL.
FT CARBOHYD 1436 1436 POTENTIAL.
FT CARBOHYD 1517 1517 POTENTIAL.
SQ SEQUENCE 1630 AA; 187289 MW; DD2F8628 CRC32;

Query Match 54.8%; Score 57; DB 1; Length 1630;
Best Local Similarity 46.2%; Pred. No. 1.53e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1464 FKKNNNEVGIADL 1476
QY 3 IKANSKFIGITEL 15

RESULT 12
ID MSP1_PLAFW STANDARD; PRT; 1639 AA.
AC P04933;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)
(PMMSA) (P195).
GN MSP-1.
OS PLASMODIUM FALCIPARUM (ISOLATE WELLCOME).
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86014355.
RA HOLDER A.A., LOCKYER M.J., ODINK K.G., SANDHU J.S., RIVEROS-MORENO V.,
RA NICHOLLS S.C., HILLMAN Y., DAVEY L.S., TIZARD M.L.V., SCHWARZ R.T.,
RA FREEMAN R.R.;
RT "Primary structure of the precursor to the three major surface
RT antigens of Plasmodium falciparum merozoites.";
RL NATURE 317:270-273(1985).
RN [2]
RP REVISIONS.
RA HOLDER A.A.;
RL SUBMITTED (MAR-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
(POTENTIAL).
CC -!- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KD, 42
CC KD AND 19 KD ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X02919; G9865; -.
DR PIR; A24594; A24594.
DR PFAM; PF00008; EGF; 1.
KW MALARIA; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
KW TRANSMEMBRANE; GPI-ANCHOR.

FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1639 MEROZOITE SURFACE PROTEIN 1.
FT CARBOHYD 116 116 POTENTIAL.
FT CARBOHYD 268 268 POTENTIAL.
FT CARBOHYD 764 764 POTENTIAL.
FT CARBOHYD 768 768 POTENTIAL.
FT CARBOHYD 783 783 POTENTIAL.
FT CARBOHYD 844 844 POTENTIAL.
FT CARBOHYD 920 920 POTENTIAL.
FT CARBOHYD 964 964 POTENTIAL.
FT CARBOHYD 1058 1058 POTENTIAL.
FT CARBOHYD 1165 1165 POTENTIAL.
FT CARBOHYD 1174 1174 POTENTIAL.
FT CARBOHYD 1445 1445 POTENTIAL.
FT CARBOHYD 1526 1526 POTENTIAL.
SQ SEQUENCE 1639 AA; 187618 MW; F0860D6A CRC32;

Query Match 54.8%; Score 57; DB 1; Length 1639;
Best Local Similarity 46.2%; Pred. No. 1.53e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1473 FKKNNNEVGIADL 1485
QY 3 IKANSKFIGITEL 15

RESULT 13
ID RIBC_BACSU STANDARD; PRI; 316 AA.
AC P54575; P70987;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE RIBOFLAVIN KINASE (EC 2.7.1.26) (FLAVOKINASE) / FMN
DE ADENYLYLTRANSFERASE (EC 2.7.7.2) (FAD PYROPHOSPHORYLASE) (FAD
DE SYNTHETASE).
GN RIBC.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA GUSAROV I.V., YOMANTAS Y.I., KOZLOV Y.I., KRENEVA R.A.,
RA PERUMOV D.A.;
RL SUBMITTED (JAN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA COQUARD D., HUECAS M., OTT M., VAN DIJL J., VAN LOON A., HOHMANN H.;
RL SUBMITTED (OCT-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: ATP + RIBOFLAVIN = ADP + FMN.
CC -!- CATALYTIC ACTIVITY: ATP + FMN = DIPHOSPHATE + FAD.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X95312; E219687; -.
DR EMBL; Z80835; E269877; -.
DR EMBL; Z99112; E1185258; -.
DR SUBTILIST; BG11495; RIBC.
KW TRANSFERASE; NUCLEOTIDYLTRANSFERASE; MULTIFUNCTIONAL ENZYME.

```

FT CONFLICT 199 199 N -> G (IN REF. 2).
SQ SEQUENCE 316 AA; 35719 MW; 2BEDCAE8 CRC32;

Query Match 53.8%; Score 56; DB 1; Length 316;
Best Local Similarity 53.8%; Pred. No. 2.44e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 284 IRSERKFKIGIKEL 296
QY 3 IKANSKFIGITEL 15

RESULT 14
ID PEPE_LACLA STANDARD; PRI; 601 AA.
AC P54124;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE OLIGOENDOPEPTIDASE F (EC 3.4.24.-).
GN PEPE.
OS LACTOCOCCUS LACTIS (SUBSP. LACTIS) (STREPTOCOCCUS LACTIS).
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; STREPTOCOCCACEAE;
OC LACTOCOCCUS.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NCDO 763;
RX MEDLINE; 95096044.
RA MONNET V., NARDI M., CHOPIN A., CHOPIN M.C., GRIPON J.C.;
RT "Biochemical and genetic characterization of PepF, an oligopeptidase
RT from Lactococcus lactis.";
RL J. BIOL. CHEM. 269:32070-32076(1994).
CC -!- FUNCTION: HYDROLYZES PEPTIDES CONTAINING BETWEEN 7 AND 17 AMINO
CC ACIDS WITH A RATHER WIDE SPECIFICITY.
CC -!- COFACTOR: BINDS A ZINC ATOM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M3 (ZINC METALLOPROTEASE);
CC ALSO KNOWN AS THE THIMET OLIGOPEPTIDASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z32522; G510140; -
CC DR EMBL; Z32522; G510140; -
CC KW HYDROLASE; METALLOPROTEASE; ZINC.
CC FT METAL 387 387 ZINC (CATALYTIC) (BY SIMILARITY).
CC FT ACI_SITE 388 388 BY SIMILARITY.
CC FT METAL 391 391 ZINC (CATALYTIC) (BY SIMILARITY).
CC SQ SEQUENCE 601 AA; 69674 MW; 2AB624A3 CRC32;

Query Match 53.8%; Score 56; DB 1; Length 601;
Best Local Similarity 46.7%; Pred. No. 2.44e+00;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 284 RYIELRKKILGITDL 298
QY 1 QYIKANSKFIGITEL 15

RESULT 15
ID ACOC_CAEEL STANDARD; PRI; 887 AA.
AC Q23500;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROBABLE ACONITATE HYDRATASE, CYTOPLASMIC (EC 4.2.1.3) (CITRATE HYDRO-
DE LYASE) (ACONITASE).
GN ZK455.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
```

```

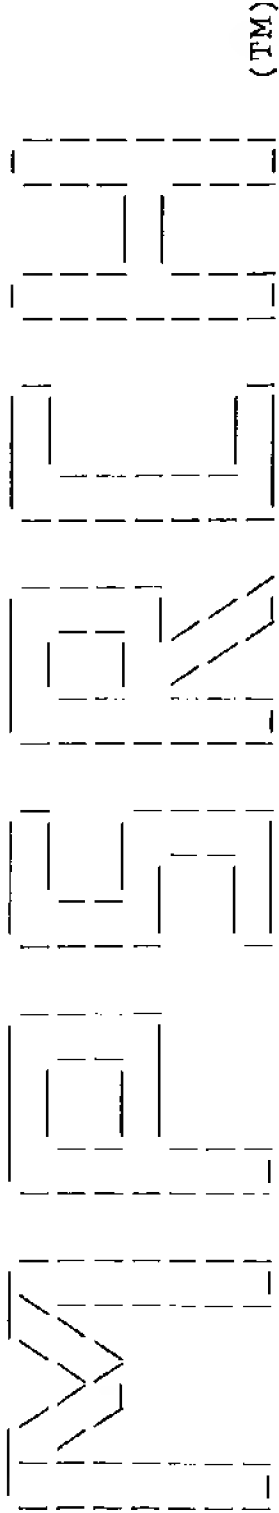
OC RHABDITINA; RHABDITIOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA WHITE S.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: CITRATE = CIS-ACONITATE + H(2)O.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z66567; E1351089; -
CC DR WORMPEP; ZK455.1; CE03812.
CC DR PROSITE; PS00450; ACONITASE_1; 1.
CC DR PROSITE; PS01244; ACONITASE_2; 1.
CC DR PFAM; PF00330; aconitase; 1.
CC DR PFAM; PF00694; Aconitase_C; 1.
CC KW LYASE; TRICARBOXYLIC ACID CYCLE; IRON-SULFUR; 4FE-4S.
CC FT METAL 436 436 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 502 502 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC FT METAL 505 505 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC SQ SEQUENCE 887 AA; 96660 MW; 226EF357 CRC32;

Query Match 53.8%; Score 56; DB 1; Length 887;
Best Local Similarity 35.7%; Pred. No. 2.44e+00;
Matches 5; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 334 QYLKSVGMFVNFTD 347
QY 1 QYIKANSKFIGITE 14

Search completed: Wed Aug 4 15:23:07 1999
Job time : 6 secs.
```

This Page Blank (uspto)



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:23:24 1999; MasPar time 7.16 seconds
Tabular output not generated. 114.393 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 26.069; Variance 32.264; scale 0.808

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	62	59.6	598	2	Q34469	2.21e-01
2	61	58.7	349	3	P78872	3.62e-01
3	59	56.7	1048	5	Q26023	9.52e-01
4	59	56.7	3119	5	Q25857	9.52e-01
5	58	55.8	131	7	Q46869	1.53e+00
6	58	55.8	357	2	Q32000	1.53e+00
7	58	55.8	357	9	Q64021	1.53e+00
8	58	55.8	423	2	Q25919	1.53e+00
9	57	54.8	400	5	Q03999	2.46e+00
10	57	54.8	421	5	Q77436	2.46e+00
11	57	54.8	424	2	Q86555	2.46e+00
12	57	54.8	539	5	Q25981	2.46e+00
13	57	54.8	539	5	Q25966	2.46e+00
14	57	54.8	539	5	Q25973	2.46e+00
15	57	54.8	539	5	Q25971	2.46e+00
16	57	54.8	539	5	Q25976	2.46e+00
17	57	54.8	539	5	Q25972	2.46e+00
18	57	54.8	539	5	Q25984	2.46e+00
19	57	54.8	1333	5	Q24262	2.46e+00
20	56	53.8	227	5	Q25803	3.91e+00

21	56	53.8	271	9	O48471	COMPLETE NUCLEOTIDE SE	3.91e+00
22	56	53.8	381	1	O59512	381AA LONG HYPOTHETICA	3.91e+00
23	56	53.8	601	2	P94880	OLIGOPEPTIDASE.	3.91e+00
24	56	53.8	1018	5	O17874	F46F6.2 PROTEIN.	3.91e+00
25	55	52.9	131	7	O46866	MHC CLASS II BETA CHAI	6.19e+00
26	55	52.9	249	7	P79565	MHC CLASS II BETA CHAI	6.19e+00
27	55	52.9	290	2	Q50869	CHEMOTACTIC RESPONSE R	6.19e+00
28	55	52.9	382	2	P94415	RESPONSE-REGULATOR ASP	6.19e+00
29	55	52.9	436	5	P91071	COSMID C18B10.	6.19e+00
30	55	52.9	447	11	P70680	ALPHA-1,3-MANNOSYL-GLY	6.19e+00
31	55	52.9	458	5	O44506	F42G8.9 PROTEIN.	6.19e+00
32	55	52.9	509	5	Q27482	SIMILAR TO CYTOCHROME	6.19e+00
33	54	51.9	108	2	Q54172	POTATIVE CYCLASE.	9.74e+00
34	54	51.9	345	2	Q30883	ERPX PROTEIN.	9.74e+00
35	54	51.9	413	3	P78759	FISSION YEAST (FRAGMEN	9.74e+00
36	54	51.9	446	10	P93632	GLOSSY15.	9.74e+00
37	54	51.9	509	5	Q27499	SIMILAR TO CYTOCHROME	9.74e+00
38	54	51.9	1102	10	O49552	UV-DAMAGED DNA-BINDING	9.74e+00
39	53	51.0	130	1	O59432	130AA LONG HYPOTHETICA	1.52e+01
40	53	51.0	188	2	Q44863	ORFB (ORF2).	1.52e+01
41	53	51.0	188	2	Q25607	HYPOTHETICAL 21.2 KD P	1.52e+01
42	53	51.0	188	2	Q45022	REPEATED DNA ELEMENT,	1.52e+01
43	53	51.0	188	2	Q44870	PLASMID, ORFA, B, C, D	1.52e+01
44	53	51.0	294	2	O25678	ATP-BINDING PROTEIN (Y	1.52e+01
45	53	51.0	652	2	O68071	DNA GYRASE SUBUNIT B (1.52e+01

ALIGNMENTS

RESULT	1
ID	O34469
AC	O34469;
DT	01-JAN-1998 (TREMBLREL. 05, CREATED)
DT	01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE	YEEB PROTEIN.
GN	YEEB.
OS	BACILLUS SUBTILIS.
OC	BACTERIA; FIRMITUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC	BACILLUS.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=168;
RX	MEDLINE; 98044033.
RA	KUNST F., OGASAWARA N., MOSZER I., ALBERTINI A.M., ALLONI G.,
RA	AZEVEDO V., BERTERO M.G., BESSIERES P., BOLOTIN A., BORCHERT S.,
RA	BORRIS R., BOURSIER L., BRANS A., BRAUN M., BRIGNELL S.C., BRON S.,
RA	BROUILLET S., BRUSCHI C.V., CALDWELL B., CAPUANO V., CARTER N.M.,
RA	CHOI S.K., CODANI J.J., CONNERTON I.F., CUMMINGS N.J., DANIEL R.A.,
RA	DENIZOT F., DEVINE K.M., DUSTERHOFT A., EHRLICH S.D., EMMERSON P.T.,
RA	ENTIAN K.D., ERRINGTON J., FABRET C., FERRARI E., FOULGER D.,
RA	FRITZ C., FUJITA M., FUJITA Y., FUMA S., GALIZZI A., GALLERON N.,
RA	GHIM S.Y., GLASER P., GOLFEAU A., GOLIGHTLY E.J., GRANDI G.,
RA	GUISEPPI G., GUY B.J., HAGA K., HAIECH J., HARWOOD C.R., HENAUT A.,
RA	HILBERT H., HOLSAPPEL S., HOSONO S., HULLO M.F., ITAYA M., JONES L.,
RA	JORIS B., KARAMATA D., KASAHARA Y., KLAERR-BLANCHARD M., KLEIN C.,
RA	KOBAYASHI Y., KOETTER P., KONINGSSTEIN G., KROGH S., KUMANO M.,
RA	KURITA K., LAPIDUS A., LARDINOIS S., LAUBER J., LAZAREVIC V.,
RA	LEE S.M., LEVINE A., LIU H., MASUDA S., MAUEL C., MEDIGUE C.,
RA	MEDINA N., MELLADO R.P., MIZUNO M., MOESTL D., NAKAI S., NOBACK M.,
RA	NOONE D., O'REILLY M., OGAWA K., OGIWARA A., OUDEGA B., PARK S.H.,
RA	PARRO V., POHL T.M., PORTELELE D., PORWOLLIK S., PRESCOTTI A.M.,
RA	PRESECAN E., PUJIC P., PURNELLE B., RAPOPORT G., REY M., REYNOLDS S.,
RA	RIEGER M., RIVOLTA C., ROCHA E., ROCHE B., ROSE M., SADAIE Y.,
RA	SATO T., SCANLAN E., SCHLEICH S., SCHROETER R., SCOFFONE F.,
RA	SEKIGUCHI J., SEKOWSKA A., SEROR S.J., SERROR P., SHIN B.S., SOLDI B.,
RA	SOROKIN A., TACCONI E., TAKAGI T., TAKAHASHI H., TAKEMARU K.,
RA	TAKUCHI M., TAMAKOSHI A., TANAKA T., TERPSTRA P., TOGNONI A.,
RA	TOSATO V., UCHIYAMA S., VANDENBOL M., VANNIER F., VASSAROTTI A.,
RA	VIARI A., WAMBUIT R., WEDLER E., WEDLER H., WEITZENEGGER T.,
RA	WINTERS P., WIPAT A., YAMAMOTO H., YAMANE K., YASUMOTO K., YATA K.,
RA	YOSHIDA K., YOSHIKAWA H.F., ZUMSTEIN E., YOSHIKAWA H., DANCHIN A.,
RT	"The complete genome sequence of the gram-positive bacterium Bacillus

CC GPI-ANCHOR.
 DR EMBL; M64681; G160539; -.
 DR PFAM; PF00008; EGF; 1.
 KW MALARIA; MEROZOITE; POLYPROTEIN; REPEAT; SIGNAL; GLYCOPROTEIN;
 KW TRANSMEMBRANE; GPI-ANCHOR.
 FT NON_TER 1 1
 FT SIGNAL <1 24
 FT CHAIN 25 400
 FT CHAIN 25 286
 FT CHAIN 287 400
 FT TRANSMEM 383 400
 SQ SEQUENCE 400 AA; 45824 MW; OE131D8C CRC32;

Query Match 54.8%; Score 57; DB 5; Length 400;
 Best Local Similarity 46.2%; Pred. No. 2.46e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 234 FKNNNEVGIADL 246
 QY 3 IKANSKFIGITEL 15

RESULT 10
 ID 077436 PRELIMINARY; PRI; 421 AA.
 AC 077436;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE EG:34F3.5 PROTEIN.
 GN EG:34F3.5.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
 OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
 OC DROSOPHILIDAE; DROSOPHILA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA CATHERINE SALLES, PHILIPPE VALENTI, ARETI DARLAMITSOU,
 RA NADINE HENDERSON, LORNA CAMPBELL, DAVID GLOVER;
 RT "Sequencing the distal x chromosome of Drosophila melanogaster."
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BENOS P.;
 RL SUBMITTED (OCT-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AL031583; E1321021; -.
 SQ SEQUENCE 421 AA; 48676 MW; 256FB6CF CRC32;

Query Match 54.8%; Score 57; DB 5; Length 421;
 Best Local Similarity 42.9%; Pred. No. 2.46e+00;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Db 198 YRTNERFINFMTL 211
 QY 2 YIKANSKFIGITEL 15

RESULT 11
 ID 086555 PRELIMINARY; PRI; 424 AA.
 AC 086555;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 45.1 KD PROTEIN.
 GN SC1F2.18.
 OS STREPTOMYCES COELICOLOR.
 OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
 OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA SEEGER K.J., HARRIS D.;
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 RX MEDLINE; 97000351.
 RA REDENBACH M., KIESER H.M., DENAPAITTE D., EICHNER A., CULLUM J.,
 RA KINASHI H., HOPWOOD D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 the 8 Mb streptomyces coelicolor A3(2) chromosome.";
 RL MOL. MICROBIOL. 21:77-96(1996).
 DR EMBL; AL031350; E1316910; -.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 424 AA; 45064 MW; B4BC2A68 CRC32;

Query Match 54.8%; Score 57; DB 2; Length 424;
 Best Local Similarity 46.7%; Pred. No. 2.46e+00;
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Db 358 QYVKAGDTLYGATDL 372
 QY 1 QYIKANSKFIGITEL 15

RESULT 12
 ID Q25981 PRELIMINARY; PRI; 539 AA.
 AC Q25981;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS PLASMODIUM FALCIPARUM.
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93295445.
 RA JONGWUTIWES S., TANABE K., KANBARA H.;
 RT "Sequence conservation in the C-terminal part of the precursor to the
 major merozoite surface proteins (MSP1) of Plasmodium falciparum from
 field isolates."
 RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
 DR EMBL; D13363; D1003128; -.
 DR PFAM; PF00008; EGF; 1.
 KW MEROZOITE; EGF-LIKE DOMAIN.
 FT NON_TER 1 1
 SQ SEQUENCE 539 AA; 61046 MW; 398440E6 CRC32;

Query Match 54.8%; Score 57; DB 5; Length 539;
 Best Local Similarity 46.2%; Pred. No. 2.46e+00;
 Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKNNNEVGIADL 385
 QY 3 IKANSKFIGITEL 15

RESULT 13
 ID Q25966 PRELIMINARY; PRI; 539 AA.
 AC Q25966;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
 DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
 GN MSP1.
 OS PLASMODIUM FALCIPARUM.
 OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93295445.
 RA JONGWUTIWES S., TANABE K., KANBARA H.;

```
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates.";
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
DR EMBL; D13357; D1003122; -.
DR PFAM; PF00008; EGF; 1.
KW MEROZOITE; EGF-LIKE DOMAIN.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61145 MW; DD8BB38E CRC32;

Query Match 54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKKNNNFVGIADL 385
:| |::|::|
QY 3 IKANSKFIGITEL 15

RESULT 14
ID Q25973 PRELIMINARY; PRT; 539 AA.
AC Q25973;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPl.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93295445.
RA JONGWUTIWES S., TANABE K., KANBARA H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates.";
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
DR EMBL; D13361; D1003126; -.
DR PFAM; PF00008; EGF; 1.
KW MEROZOITE; EGF-LIKE DOMAIN.
FT NON_TER 1
SQ SEQUENCE 539 AA; 61047 MW; D7140867 CRC32;

Query Match 54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKKNNNFVGIADL 385
:| |::|::|
QY 3 IKANSKFIGITEL 15

RESULT 15
ID Q25971 PRELIMINARY; PRT; 539 AA.
AC Q25971;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE MAJOR MEROZOITE SURFACE PROTEIN PRECURSOR (FRAGMENT).
GN MSPl.
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93295445.
RA JONGWUTIWES S., TANABE K., KANBARA H.;
RT "Sequence conservation in the C-terminal part of the precursor to the
RT major merozoite surface proteins (MSPl) of Plasmodium falciparum from
RT field isolates.";
RL MOL. BIOCHEM. PARASITOL. 59:95-100(1993).
DR EMBL; D13359; D1003124; -.
DR PFAM; PF00008; EGF; 1.
KW MEROZOITE; EGF-LIKE DOMAIN.
```

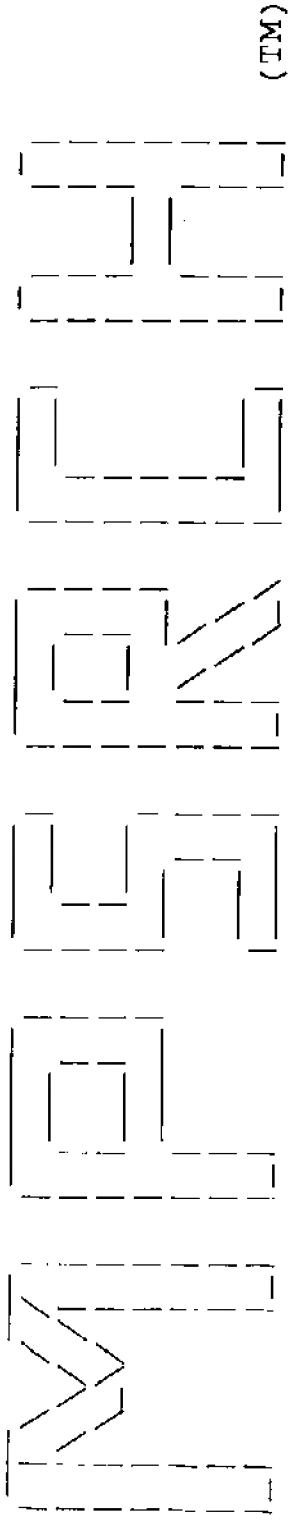
```
FT NON_TER 1
SQ SEQUENCE 539 AA; 61017 MW; 7A8981F1 CRC32;

Query Match 54.8%; Score 57; DB 5; Length 539;
Best Local Similarity 46.2%; Pred. No. 2.46e+00;
Matches 6; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 373 FKKNNNFVGIADL 385
:| |::|::|
QY 3 IKANSKFIGITEL 15

Search completed: Wed Aug 4 15:23:35 1999
Job time : 11 secs.
```

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:44:45 1999; MasPar time 3.52 Seconds
Tabular output not generated. 90.550 Million cell updates/sec

Title: >US-09-049-847-1
Description: (1-15) from US09049847.pep
Perfect Score: 104
Sequence: 1 QYIKANSKFIGITEL 15

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 18.668; Variance 53.652; scale 0.348

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	104	100.0	15	38	W67578 T-cell epitope peptid	4.95e-05
2	104	100.0	15	23	W11505 Tetanus toxoid univer	4.95e-05
3	104	100.0	15	27	W35506 Universal T-cell epit	4.95e-05
4	104	100.0	15	1	W06310 Tetanus toxin epitope	4.95e-05
5	104	100.0	15	37	W73220 Tetanus toxoid epitop	4.95e-05
6	104	100.0	15	35	W71321 Universal helper T-ce	4.95e-05
7	104	100.0	15	36	W67033 Tetanus toxin fragmen	4.95e-05
8	104	100.0	16	27	W35445 T-cell stimulatory pe	4.95e-05
9	104	100.0	17	16	R82573 Tetanus toxin helper	4.95e-05
10	104	100.0	17	19	W05599 Tetanus toxin helper	4.95e-05
11	104	100.0	17	13	R62692 Helper T cell epitope	4.95e-05
12	104	100.0	17	17	R88395 T-cell antigen TT2 pe	4.95e-05
13	104	100.0	27	16	R82596 IgE CH4 region contg.	4.95e-05
14	104	100.0	27	13	R62701 LHRH-containing immun	4.95e-05
15	104	100.0	29	16	R83561 IgE CH4 region contg.	4.95e-05
16	104	100.0	30	10	R44398 HIV antigen fragment.	4.95e-05

17	104	100.0	37	13	R65383	Universal immunostimu	4.95e-05
18	104	100.0	37	13	R65389	Universal immunostimu	4.95e-05
19	104	100.0	47	13	R62723	LHRH-containing immun	4.95e-05
20	104	100.0	50	20	W06131	Anti-cholesteryl este	4.95e-05
21	104	100.0	573	2	P70345	Portion of B fragment	4.95e-05
22	97	93.3	31	20	W06129	Anti-cholesteryl este	3.91e-04
23	97	93.3	63	3	R14263	Immunogenic branched	3.91e-04
24	96	92.3	14	8	R46509	Tetanus toxoid residu	5.23e-04
25	96	92.3	14	15	R78918	Tetanus toxoid 830-84	5.23e-04
26	96	92.3	14	20	W03003	Carrier peptide for a	5.23e-04
27	96	92.3	14	27	W35437	T-cell stimulatory pe	5.23e-04
28	96	92.3	14	15	R75943	T helper epitope from	5.23e-04
29	96	92.3	14	13	R70910	Tetanus toxoid 830-84	5.23e-04
30	96	92.3	14	30	W50108	Pan DR binding peptid	5.23e-04
31	96	92.3	14	1	R06309	Tetanus toxin epitope	5.23e-04
32	96	92.3	14	6	R33497	T helper peptide teta	5.23e-04
33	96	92.3	14	26	R74160	Antigenic peptide TT	5.23e-04
34	96	92.3	15	23	W11506	Tetanus toxoid mutant	5.23e-04
35	96	92.3	15	37	W73221	Tetanus toxoid epitop	5.23e-04
36	96	92.3	27	15	R78713	HBV specific cytotoxi	5.23e-04
37	96	92.3	27	6	R33507	T helper epitope/HBV	5.23e-04
38	96	92.3	29	33	W48992	Lipidated vaccine 2 a	5.23e-04
39	96	92.3	29	33	W48991	Lipidated vaccine 1 a	5.23e-04
40	96	92.3	30	23	W08413	Synthetic lipopeptide	5.23e-04
41	96	92.3	50	28	W46447	CETP B cell epitope/L	5.23e-04
42	96	92.3	50	20	W06132	Anti-cholesteryl este	5.23e-04
43	94	90.4	14	26	R74167	TT 830-843 based anti	9.39e-04
44	93	89.4	19	35	W78830	Tetanus toxoid protei	1.26e-03
45	91	87.5	14	1	R06318	Tetanus toxin epitope	2.24e-03

ALIGNMENTS

RESULT 1
ID W67578 standard; peptide; 15 AA.

AC W67578; 99-044514/04.
DT 02-MAR-1999 (first entry)
DE T-cell epitope peptide #4 for chimeric fimbria/T-cell epitope peptide.
KW Chimeric; non-typable Haemophilus influenzae; fimbria; T-cell epitope; immunogenic composition; immune response.
OS Synthetic.
PN US5843464-A.
PD 01-DEC-1998.
PF 02-JUN-1995; 460502.
PR 02-JUN-1995; US-460502.
PA (OHIS) UNIV OHIO STATE.
PI Bakaletz LO, Kaumaya PTP;
DR WPI; 99-044514/04.
PT Synthetic chimeric fimbria peptide - useful for vaccination against non-typable Haemophilus influenzae
PS Disclosure; Column 4; 16pp; English.
CC The invention relates to the manufacture of a synthetic chimeric peptide comprising a non-typable Haemophilus influenzae fimbria peptide fused via a linker peptide to a T-cell epitope peptide. The chimeric peptide is used in immunogenic compositions which induce an immune response against non-typable Haemophilus influenzae. This sequence represents an example of a T-cell epitope peptide used to generate the chimeric peptide.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 38; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
QY 1 QYIKANSKFIGITEL 15

RESULT 2
ID W11505 standard; Protein; 15 AA.
AC W11505;
DT 24-SEP-1997 (first entry)
DE Tetanus toxoid universal Th epitope TT830.

KW Humanised antibody; anti-Fc receptor; H22; bifunctional; bispecific;
KW fusion protein; chimera; tetanus toxoid; helper T cell epitope;
KW antigen presentation; ds.
OS Clostridium tetani.
PN WO9640789-A1.
PD 19-DEC-1996.
PF 07-JUN-1996; U09988.
PR 07-JUN-1995; US-484172.
PA (MEDA-) MEDAREX INC.
PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
DR WPI; 97-052242/05.
DR N-PSDB; T58127.
PT Recombinant, multi-specific anti-Fc receptor antibody molecules -
PT also comprise an anti-target portion, used for the treatment of
PT cancer, autoimmune disease and pathogenic infection
PS Example 7; Fig 24; 115pp; English.
CC Synthetic DNA coding for the wild-type universal Th epitope from
CC tetanus toxoid, designated Tt830, was fused to the 3'-end of DNA
CC encoding heavy chain sequences from the humanised anti-Fc gamma RI
CC monoclonal antibody H22. The resulting fusion protein was shown to
CC be significantly more efficient in antigen presentation and T cell
CC stimulation that the Tt830 epitope alone. A similar fusion
CC construct was prepared coding for a mutant, antagonistic form of the
CC epitope (designated Tt833S) fused to the anti-Fc gamma RI. The
CC Fab22-Tt833S is at least 100 times more effective than Tt833S in
CC inhibiting T cell activation.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 23; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 3

ID W35506 standard; peptide; 15 AA.
AC W35506;
DT 22-APR-1998 (first entry)
DE Universal T-cell epitope peptide SEQ ID NO:8.
KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
KW scaffold; inhibition; metastasis; wound healing; solid phase.
OS Unidentified.
PN WC9738011-A1.
PD 16-OCT-1997.
PF 03-APR-1997; D00146.
PR 03-APR-1996; DK-000398.
PA (PEPR-) PEPRESEARCH AS.
PI Heegaard PMH, Jakobsen PH;
DR WPI; 97-512645/47.
PT Non-dendritic peptide carrier linked to a solid phase - useful as a
PT diagnostic agent and as a scaffold for production of chemical
PT derivatives
PS Example 20; Page 124; 262pp; English.
CC A non-dendritic peptide carrier (A) has been developed which is coupled
CC through a linker to a solid phase, forming a complex of (A)-solid phase.
CC Where (A) comprises 10-50 amino acids capable of forming a secondary
CC structure in a benign buffer after liberation from the solid phase, and
CC further the (A)-solid phase complex comprises an immunogenic substance
CC and/or an immune mediator coupled on (A). The present sequence
CC represents a peptide used in an example from the present invention. An
CC (A)-solid phase complex can be used as a scaffold for the production of
CC chemical derivatives, characterised by covalently attaching molecules at
CC attachment points. Alternatively (A) is used as a scaffold-peptide for
CC the incorporation into an Immunostimulating Complex (Iscom) resulting an
CC (A)-Iscom complex which is used for the chemical coupling of antigenic
CC substances in an aqueous solution by conjugation. (A) derivatised with
CC one or more peptides having fibronectin-, laminin- or vitronectin-like
CC binding activities can be used for the promotion of cell-attachment to
CC plastic surfaces, in particular to inhibit tumour growth and metastasis,
CC and for promotion of wound healing. Also a derivatised (A) can be used

CC for the selection of specifically-binding aptamers or as a diagnostic
CC agent. Such diagnostic-(A) molecules could be used to detect molecules
CC derived from or indicative of pregnancy or of a disease, such as an
CC infectious, autoimmune or cancerous disease.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 27; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 4

ID R06310 standard; protein; 15 AA.
AC R06310;
DT 04-DEC-1990 (first entry)
DE Tetanus toxin epitope.
KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
KW antimalarial.
OS Synthetic.
PN EP-378881-A.
PD 25-JUL-1990.
PF 27-DEC-1989; 203318.
PR 17-JAN-1989; IT-019110.
PR 16-NOV-1989; IT-022409.
PA (ENIE) ENRICERCE SPA.
PI Pessi A, Bianchi E, Verdini AS, Corradin G;
DR WPI; 90-225582/30.
PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
PT as universal carriers for prepn. of immunogenic conjugate(s) for
PT use as vaccines.
PS Claim 1; Page 17; 20pp; English.
CC Epitopic peptides may be used with synthetic hapten derived from
CC a pathogen to generate an immune response to the pathogen.
CC Peptides are recognised by numerous T-helper cell clones within
CC the context of a wide range of alleles of the human MHC.
CC The peptides may be used in an antimalarial vaccine inducing Ab.
CC response to P.falciparum.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 5

ID W73220 standard; Protein; 15 AA.
AC W73220;
DT 25-JAN-1999 (first entry)
DE Tetanus toxoid epitope.
KW Multispecific single chain antibody; antibody H22; tumour cell; therapy;
KW antibody-dependent cellular cytotoxicity; ADCC; HER 2/neu; infection;
KW epidermal growth factor receptor; breast cancer; ovarian cancer.
OS Synthetic.
PN US5837243-A.
PD 17-NOV-1998.
PF 07-JUN-1996; 661052.
PR 07-JUN-1996; US-661052.
PR 07-JUN-1995; US-484172.
PA (MEDA-) MEDAREX INC.
PI Deo YM, Goldstein J, Graziano R, Somasundaram C;
DR WPI; 99-023374/02.
PT Specific killing of tumour cells - using a multi-specific molecule
PT comprising an anti-Fc receptor antibody and a portion which binds to
PT a target cell
PS Example 7; Column 27; 57pp; English.

CC This sequence represents a tetanus toxoid epitope and is recognised
CC by the multispecific single chain antibody designated H22. The
CC antibody can be used in the method of the invention for inducing
CC antibody-dependent cellular cytotoxicity (ADCC) against a tumour cell
CC which is characterised by overexpression of HER 2/neu or epidermal growth
CC factor receptor (EGFR), comprises contacting the tumour cell with a
CC multispecific protein molecule (preferably a single chain antibody)
CC comprising: (a) an anti-Fc receptor antibody or an antigen binding
CC fragment; (b) a portion which binds to HER 2/neu; and (c) a portion which
CC binds to EGFR. The method can be used for treating cancers especially
CC breast cancer or ovarian cancer. The multispecific antibody can also
CC be administered prophylactically to vaccinate a subject against infection
CC by a target cell.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 37; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 6

ID W71321 standard; peptide; 15 AA.
AC W71321;
DT 26-NOV-1998 (first entry)
DE Universal helper T-cell epitope P2 derived from tetanus toxin.
KW Liver stage; Plasmodium; Navy Yoelii Liver Stage 3 antibody; NYLS3;
KW hepatic and erythrocytic stage protein; PyHEP17; vaccine;
KW malaria parasite; teanus toxin; P2; helper T-cell epitope.
OS Synthetic.
OS Clostridium tetani.
PN US5814617-A.
PD 29-SEP-1998.
PF 07-OCT-1994; 319704.
PR 07-OCT-1994; US-319704.
PA (USNA) US SEC OF NAVY.
PI Charoenvit Y, Doolan DL, Hedstrom RC, Hoffman SL;
DR WPI; 98-541794/46.
PT Vaccine for protecting mammal against infection by malaria caused by
PT Plasmodium species - comprises a first nucleic acid encoding a first
PT polypeptide capable of eliciting an immune reaction against an
PT antigen expressed during the liver
PS Disclosure; Column 12; 24pp; English.
CC W71321-22 represent universal helper T-cell epitopes derived from
CC tetanus toxin. They are used to enhance host immune response to
CC vaccines. The specification describes a Plasmodium yoelii liver stage
CC 17 kDa hepatic and erythrocytic stage protein designated PyHEP17. This
CC protein elicits a response from an Igl monoclonal antibody designated
CC Navy Yoelii Liver Stage 3 (NYLS3). This antibody does not recognise
CC sporozoites, but does recognise P. yoelii liver stage parasites. NYLS3
CC eliminates upto 90% of liver stage parasites. The specification describes
CC a vaccine for reducing the severity or incidence of infection by a
CC malaria parasite of the genus Plasmodium. The DNA vaccine comprises
CC exon 1 and part of exon 2 of the PyHEP17 gene.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 35; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 7

ID W67033 standard; peptide; 15 AA.
AC W67033;
DT 15-DEC-1998 (first entry)
DE Tetanus toxin fragment (residues 830-844).

KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KW dendrimeric poly-lysine; epitope; tumour.
OS Clostridium tetani.
PN W09843677-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; E01922.
PR 27-MAR-1997; US-041726.
PA (INSP) INST PASTEUR.
PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;
DR WPI; 98-557071/47.
PT Carbohydrate peptide conjugate used as vaccine - comprises carrier
PT with dendrimeric poly-lysine enabling multiple epitopes to be
PT covalently attached
PS Disclosure; Page 13; 55pp; English.
CC The invention relates to a new carbohydrate peptide conjugate, which
CC comprises a carrier with a dendrimeric poly-lysine enabling multiple
CC epitopes to be covalently attached to it. Also claimed are: (1) an
CC antibody purified from biological fluid or cells of organisms
CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis
CC kit comprising antigen-specific antibodies elicited by immunisation with
CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and
CC diagnosis kit are used to provide pharmaceutical compositions and
CC vaccines against tumours. These can be used to support an immune response
CC against viral infections caused by hepatitis virus, HIV or cytomegalo
CC virus. They can be used to enhance immune responses, especially B- and T-
CC cell responses, of humans and animals against bacterial infections. The
CC carbohydrate peptide conjugate stimulates the antibody and T-cell
CC response without stimulating undesired immune responses. The composition
CC is capable of increasing the survival of tumour bearing humans and
CC animals. The present sequence corresponds to residues 830-844 of tetanus
CC toxin. The synthetic peptide corresponding to this sequence may be used
CC as an epitope in a carbohydrate peptide conjugate.
SQ Sequence 15 AA;

Query Match 100.0%; Score 104; DB 36; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qvikanskfigitel 15
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 8

ID W35445 standard; peptide; 16 AA.
AC W35445;
DT 22-APR-1998 (first entry)
DE T-cell stimulatory peptide SEQ ID NO:51.
KW T-cell stimulatory peptide; immunogen; non-dendritic; carrier; tumour;
KW scaffold; inhibition; metastasis; wound healing; solid phase.
OS Unidentified.
PN W09738011-A1.
PD 16-OCT-1997.
PF 03-APR-1997; D00146.
PR 03-APR-1996; DK-000398.
PA (PEPR-) PEPRESEARCH AS.
PI Heegaard PMH, Jakobsen PH;
DR WPI; 97-512645/47.
PT Non-dendritic peptide carrier linked to a solid phase - useful as a
PT diagnostic agent and as a scaffold for production of chemical
PT derivatives
PS Claim 30; Page 199; 262pp; English.
CC A non-dendritic peptide carrier (A) has been developed which is coupled
CC through a linker to a solid phase, forming a complex of (A)-solid phase.
CC Where (A) comprises 10-50 amino acids capable of forming a secondary
CC structure in a benign buffer after liberation from the solid phase, and
CC further the (A)-solid phase complex comprises an immunogenic substance
CC and/or an immune mediator coupled on (A). The present sequence
CC represents a specifically claimed T-cell stimulatory peptide from the
CC present invention. An (A)-solid phase complex can be used as a scaffold
CC for the production of chemical derivatives, characterised by covalently
CC attaching molecules at attachment points. Alternatively (A) is used as
CC a scaffold-peptide for the incorporation into an Immunostimulating

CC Complex (Iscom) resulting an (A)-Iscom complex which is used for the
CC chemical coupling of antigenic substances in an aqueous solution by
CC conjugation. (A) derivatised with one or more peptides having
CC fibronectin-, laminin- or vitronectin-like binding activities can be
CC used for the promotion of cell-attachment to plastic surfaces, in
CC particular to inhibit tumour growth and metastasis, and for promotion
CC of wound healing. Also a derivatised (A) can be used for the selection
CC of specifically-binding aptamers or as a diagnostic agent. Such
CC diagnostic-(A) molecules could be used to detect molecules derived from
CC or indicative of pregnancy or of a disease, such as an infectious,
CC autoimmune or cancerous disease.
SQ Sequence 16 AA;

Query Match 100.0%; Score 104; DB 27; Length 16;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
| | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 9

ID R82573 standard; peptide; 17 AA.
AC R82573;
DT 13-JUN-1996 (first entry)
DE Tetanus toxin helper T cell epitope, TT1.
KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
KW vaccine; allergy; antibody; constant heavy chain.
OS Clostridium tetani.
PN W09526365-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; U03741.
PR 28-MAR-1994; US-218461.
PR 25-OCT-1994; US-328912.
PA (UNBI-) UNITED BIOMEDICAL INC.
PI Wang CY;
DR WPI; 95-351297/45.
PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
PT T cell epitope - useful for eliciting antibody prodn. for allergy
PT treatment
PS Claim 3; Page 59; 87pp; English.
CC R82571-91 are helper T cell epitopes which can be used in the
CC preparation of a peptide immunogen that is useful in vaccines for
CC treating allergic reactions. In the immunogen an IgE CH4 peptide
CC is attached C-terminally to a series of amino acids including a
CC helper T cell epitope. The immunogen may also opt. contain a fatty
CC acid or fatty acid derivative, an invasin domain or alpha-NH2. The
CC immunogen produces high titres of antibodies to the effector site
CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
CC cell activation and reduce allergen-induced IgE prodn. The immunogens
CC may be used in either a radially branching multimeric form or a
CC linearly arranged monomeric form.
SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 16; Length 17;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qyikanskfigitel 17
| | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 10

ID W05599 standard; peptide; 17 AA.
AC W05599;
DT 10-DEC-1996 (first entry)
DE Tetanus toxin helper T cell epitope #1.
KW Immunoglobulin; IgE; membrane protein; human; epsilon chain; hepatitis B;
KW membrane anchoring domain; helper T cell; surface antigen; core antigen;
KW pertussis toxin; tetanus toxin; measles virus F protein; immunotherapy;
KW Chlamydia trachomatis major outer membrane protein; immunogen; vaccine;

KW diphtheria toxin; plasmodium falciparum; circumsporozoite; E. coli Trat;
KW schistosoma mansoni; triose phosphate isomerase; allergenic reaction;
KW allergic rhinitis; food allergy; anaphylaxis; virally-induced asthma;
KW antihistamine; decongestant; beta-2 agonist; immunosuppression;
KW corticosteroid.
OS Synthetic.
PN W09612740-A1.
PD 02-MAY-1996.
PF 25-OCT-1995; U13841.
PR 25-OCT-1994; US-328519.
PR (UNBI-) UNITED BIOMEDICAL INC.
PI Walfield AM, Wang CY;
DR WPI; 96-230555/23.
PT Peptide immunogen useful in treatment of allergy - comprises
PT membrane-bound IgE epsilon-chain peptide synthesised linearly in
PT tandem with T helper epitope peptide
PS Claim 2; Page 18; 53pp; English.
CC W05957-W05616 represent helper T cell epitopes used in the peptide
CC immunogens of the invention. This sequence represents the tetanus toxin
CC helper T cell antigen. The peptides of the invention contain one of
CC these sequences, and a membrane-bound immunoglobulin E (IgE) fragment
CC (see W05595 and W05596). The peptide immunogens of the invention can be
CC used in vaccines for the immunotherapeutic treatment of allergenic
CC reactions, including allergic rhinitis, food allergies, anaphylaxis, or
CC virally-induced asthma. The immunogens overcome the short effective
CC period of antihistamines, decongestants, and beta-2 agonists, while
CC preventing the broad immunosuppression of corticosteroids. The peptides
CC do not have the potential side effects of restlessness or sedation
CC (associated with antihistamines), associated increased morbidity in
CC asthmatics (as seen with beta-2 agonists) and adverse hormonal activities
CC (observed in corticosteroid users).
SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 19; Length 17;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qyikanskfigitel 17
| | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 11

ID R62692 standard; peptide; 17 AA.
AC R62692;
DT 10-SEP-1995 (first entry)
DE Helper T cell epitope for use in universal immune stimulator.
KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
KW vaccine; tetanus toxin.
OS Clostridium tetani.
PN W09425060-A.
PD 10-NOV-1994.
PF 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
PI Ladd AE, Wang CY, Zamb T;
DR WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PT that suppress LHRH activity in males and females
PS Claim 7; Page 25; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein haptens containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC potent immune responses to the coupled peptide or protein. The
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
CC which elicits an immune response to the coupled peptide in members of
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence from the invasin protein of Yersinia.
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
CC invasin and Th domains and between the immune stimulator and haptens

CC components. When the hapten is LHRH, then optionally the invasin domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents a tetanus toxin helper T cell epitope
 CC which can be used as Th in the immune stimulator.
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 13; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qyikanskfigitel 17
 QY 1 QYIKANSKFIGITEL 15

RESULT 12

ID R88395 standard; Peptide; 17 AA.
 AC R88395;
 DT 12-JUN-1996 (first entry)
 DE T-cell antigen TT2 peptide.
 KW T-antigen; vaccine; antibody; T-cell; T-lymphocyte;
 KW alpha-helix; coiled-coil heterodimer; core peptide; subunit.
 OS Synthetic.
 PN W09531480-A1.
 PD 23-NOV-1995.
 PF 18-MAY-1995; CA0293.
 PR 18-MAY-1994; US-245507.
 PA (SPIS-) SPI SYNTHETIC PEPTIDES INC.
 PI Cachia PJ, Hodges RS, Houston ME, Irvin RT, Kay CM;
 PI Zhou NE;
 DR WPI; 96-010880/01.
 PT Hetero:dimeric polypeptide immunogen in coiled-coil configuration
 PT with different antigens on each sub:unit - useful in vaccines and
 PT for antibody prodn.
 PS Claim 7; Page 61; 95pp; English.
 CC This T-cell antigen TT2 peptide may be attached to a core peptide
 CC contained in one of the 2 subunits of an alpha-helical coiled-coil
 CC heterodimer. Each core peptide is comprised of terminal and
 CC internal AA repeat sequences. This peptide antigen is attached
 CC to the core peptide through covalent linkages to certain AA of the
 CC internal repeats. The 2 subunits of the heterodimer are arranged
 CC in a stable alpha-helical coiled-coil configuration having a 1:1
 CC stoichiometry, and the peptide antigen is disposed toward the outer
 CC surfaces of the configuration. The heterodimer may be used as a
 CC synthetic vaccine (optionally multivalent) or to generate
 CC antibodies.
 SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 17; Length 17;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 qyikanskfigitel 15
 QY 1 QYIKANSKFIGITEL 15

RESULT 13

ID R82596 standard; peptide; 27 AA.
 AC R82596;
 DT 13-JUN-1996 (first entry)
 DE IGE CH4 region contg. peptide immunogen for treating allergies.
 KW IGE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
 KW vaccine; allergy; antibody; constant heavy chain.
 OS Synthetic.
 PN W09526365-A1.
 PD 05-OCT-1995.
 PF 24-MAR-1995; U03741.
 PR 28-MAR-1994; US-218461.
 PR 25-OCT-1994; US-328912.
 PA (UNBI-) UNITED BIOMEDICAL INC.
 PI Wang CY;
 DR WPI; 95-351297/45.

PT Synthetic peptide-based immunogen contg. IGE CH4 peptide and helper
 PT T cell epitope - useful for eliciting antibody prodn. for allergy
 PT treatment
 PS Claim 5; Page 52; 87pp; English.
 CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
 CC useful in vaccines for treating allergic reactions. In the immunogens,
 CC an IGE CH4 peptide is attached C-terminally to a series of amino acids
 CC including a helper T cell epitope. The immunogen may also opt. contain
 CC a fatty acid or fatty acid derivative, an invasin domain or alpha-NH2.
 CC The immunogen produces high titres of antibodies to the effector site
 CC in human IGE heavy chain (the CH4 domain peptide) which inhibit mast
 CC cell activation and reduce allergen-induced IGE prodn. The immunogens
 CC may be used in either a radially branching multimeric form or a
 CC linearly arranged monomeric form.
 SQ Sequence 27 AA;

Query Match 100.0%; Score 104; DB 16; Length 27;
 Best Local Similarity 100.0%; Pred. No. 4.95e-05;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 qyikanskfigitel 17
 QY 1 QYIKANSKFIGITEL 15

RESULT 14

ID R62701 standard; peptide; 27 AA.
 AC R62701;
 DT 10-SEP-1995 (first entry)
 DE LHRH-containing immunogenic peptide.
 KW Helper T cell epitope; universal immune stimulator; invasin; hapten;
 KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;
 KW androgen-dependent carcinoma; antitumour; infertility;
 OS Synthetic.
 PS Key
 FH Location/Qualifiers
 FT domain 1..17
 FT domain /note= "tetanus toxin helper T cell epitope"
 FT domain 18..27
 FT domain /note= "LHRH hapten"

WO9425060-A.
 10-NOV-1994.
 28-APR-1994; U04832.
 27-APR-1993; US-057166.
 14-APR-1994; US-229275.
 (LADD/) LADD A E.
 (WANG/) WANG C Y.
 (ZAMB/) ZAMB I.
 Ladd AE, Wang CY, Zamb T;
 WPI; 94-357910/44.
 PT Immunogenic luteinising hormone releasing hormone peptide(s) -
 PT that suppress LHRH activity in males and females
 PS Claims 8, 12; Page 84; 213pp; English.
 CC Synthetic immunogenic peptides are provided in which a universal immune
 CC stimulator is linked to a peptide or protein hapten containing B cell
 CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
 CC potent immune responses to the coupled peptide or protein. The
 CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
 CC which elicits an immune response to the coupled peptide in members of
 CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
 CC an adjuvant peptide sequence from the invasin protein of Yersinia.
 CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
 CC invasin and Th domains and between the immune stimulator and hapten
 CC components. When the hapten is LHRH, then optionally the invasin domain
 CC can be omitted from the immune stimulator component.
 CC The present sequence represents an LHRH-containing, invasin-free
 CC immunogenic peptide as above which can be used as a potent vaccine for
 CC treating e.g. prostatic hyperplasia, androgen-dependent carcinoma,
 CC prostatic carcinoma, testicular carcinoma, endometriosis, benign
 CC uterine tumours, recurrent functional ovarian cysts, (severe)
 CC premenstrual syndrome or oestrogen-dependent breast cancer, or for
 CC induction of infertility.
 CC This sequence is particularly preferred.

```
SQ      Sequence      27 AA;

Query Match          100.0%; Score 104; DB 13; Length 27;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      3 qvikanskfigitel 17
        |||||
QY      1 QYIKANSKFIGITEL 15

RESULT 15
ID R83561 standard; peptide; 29 AA.
AC R83561;
DT 13-JUN-1996 (first entry)
DE IgE CH4 region contg. peptide immunogen for treating allergies.
KW IgE; CH4; immunoglobulin; epsilon; immunogen; helper T cell; epitope;
KW vaccine; allergy; antibody; constant heavy chain.
OS Synthetic.
PN WO9526365-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; U03741.
PR 28-MAR-1994; US-218461.
PR 25-OCT-1994; US-328912.
PA (UNBI-) UNITED BIOMEDICAL INC.
PI Wang CY;
DR WPI; 95-351297/45.
PT Synthetic peptide-based immunogen contg. IgE CH4 peptide and helper
PT T cell epitope - useful for eliciting antibody prodn. for allergy
PT treatment
PS Claim 5; Page 68-69; 87pp; English.
CC R82592-R82600 and R83560-R83581 are peptide immunogens that are
CC useful in vaccines for treating allergic reactions. In the immunogens,
CC an IgE CH4 peptide is attached C-terminally to a series of amino acids
CC including a helper T cell epitope. The immunogen may also opt. contain
CC a fatty acid or fatty acid derivative, an invasin domain or alpha-NH2.
CC The immunogen produces high titres of antibodies to the effector site
CC in human IgE heavy chain (the CH4 domain peptide) which inhibit mast
CC cell activation and reduce allergen-induced IgE prodn. The immunogens
CC may be used in either a radially branching multimeric form or a
CC linearly arranged monomeric form.
SQ      Sequence      29 AA;

Query Match          100.0%; Score 104; DB 16; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      3 qvikanskfigitel 17
        |||||
QY      1 QYIKANSKFIGITEL 15

RESULT 16
ID R44398 standard; peptide; 30 AA.
AC R44398;
DT 08-NOV-1994 (first entry)
DE HIV antigen fragment.
KW HIV; human immunodeficiency virus; immunisation; monoclonal antibody.
OS Human immunodeficiency virus.
PN TW-208717-A.
PD 01-JUL-1993.
PF 24-APR-1992; 103240.
PR 24-APR-1992; TW-103240.
PA (CHIN/) CHIN L.
PI Chin L;
DR WPI; 93-335491/42.
PT Induction of neutralising human monoclonal antibodies against
PT human immuno: deficiencies - by sepg. peripheral mono:nuclear cells
PT from blood using density gradient centrifugation, and treating
PT cells by L-leucyl-L-leucine methyl ester etc.
PS Claim 1; ; 36pp; Taiwanese.
CC The invention relates to a method of assessing human
CC immunodeficiency virus and producing human immunodeficiency
```

```
CC antibodies by in-vitro immunisation, which comprises: (a) separating
CC peripheral mononuclear cells from blood using density gradient
CC centrifugation; (b) treating the mononuclear cells with L-leucyl-L-
CC leucine methyl ester; and (c) using the present antigen fragment,
CC which is formed by coupled T and B cells, in a culture medium of
CC human serum, IL-2 and T cells to effect cultivation and achieve in
CC vitro immunisation.
SQ      Sequence      30 AA;

Query Match          100.0%; Score 104; DB 10; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 qvikanskfigitel 15
        |||||
QY      1 QYIKANSKFIGITEL 15

RESULT 17
ID R65383 standard; peptide; 37 AA.
AC R65383;
DT 21-SEP-1995 (first entry)
DE Universal immunostimulator having GG spacers.
KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
KW tetanus toxin.
OS Synthetic.
FH Key
FT domain
FT /note= "invasin domain"
FT domain
FT /note= "tetanus toxin helper T cell epitope"
PN WO9425060-A.
PD 10-NOV-1994.
PF 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
PI Ladd AE; Wang CY; Zamb T;
DR WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PT that suppress LHRH activity in males and females
PS Disclosure; Page 95; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein haptens containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC potent immune responses to the coupled peptide or protein. The
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
CC which elicits an immune response to the coupled peptide in members of
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence from the invasin protein of yersinia.
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
CC invasin and Th domains and between the immune stimulator and haptens
CC components. When the haptens is LHRH, then optionally the invasin domain
CC can be omitted from the immune stimulator component.
CC The present sequence is an example of an invasin-GG-Th-GG- immune
CC stimulator to which a haptens can be bonded.
SQ      Sequence      37 AA;

Query Match          100.0%; Score 104; DB 13; Length 37;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      21 qvikanskfigitel 35
        |||||
QY      1 QYIKANSKFIGITEL 15

RESULT 18
ID R65389 standard; peptide; 37 AA.
AC R65389;
DT 21-SEP-1995 (first entry)
```


DE Universal immunostimulator having GG spacers.
KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
KW tetanus toxin.
OS Synthetic.
FH Key Location/Qualifiers
FT domain 3..19
FT domain /note= "tetanus toxin helper T cell epitope"
FT domain 22..37
FT domain /note= "invasin domain"
PN WO9425060-A.
PD 10-NOV-1994.
PF 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB T.
PI Ladd AE, Wang CY, Zamb I;
DR WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PS that suppress LHRH activity in males and females
PS Disclosure; Page 95; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein haptens containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC which elicits an immune response to the coupled peptide in members of
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence (e.g. Gly-Gly) can be provided between the
CC invasin and Th domains and between the immune stimulator and haptens
CC components. When the haptens is LHRH, then optionally the invasin domain
CC can be omitted from the immune stimulator component.
CC The present sequence represents an LHRH-containing immunogenic peptide
CC as above which can be used as a potent vaccine for treating e.g.
CC prostatic hyperplasia, androgen-dependent carcinoma, prostatic
CC carcinoma, testicular carcinoma, endometriosis, benign uterine tumours,
CC recurrent functional ovarian cysts, (severe) premenstrual syndrome or
CC oestrogen-dependent breast cancer, or for induction of infertility.
SQ Sequence 37 AA;
Query Match 100.0%; Score 104; DB 13; Length 37;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 5 qyikanskfigitel 19
QY 1 QYIKANSKFIGITEL 15
RESULT 19
ID R62723 standard; peptide; 47 AA.
AC R62723;
DT 17-SEP-1995 (first entry)
DE LHRH-containing immunogenic peptide.
KW Helper T cell epitope; universal immune stimulator; invasin; haptens;
KW vaccine; LHRH; luteinising hormone releasing hormone; prostate;
KW androgen-dependent carcinoma; antitumour; infertility;
KW tetanus toxin.
OS Synthetic.
FH Key Location/Qualifiers
FT domain 1..16
FT domain /note= "invasin domain"
FT domain 19..35
FT domain /note= "tetanus toxin helper T cell epitope"
FT domain 38..47
FT domain /note= "LHRH haptens"
PN WO9425060-A.
PD 10-NOV-1994.
PF 28-APR-1994; U04832.
PR 27-APR-1993; US-057166.
PR 14-APR-1994; US-229275.
PA (LADD/) LADD A E.
PA (WANG/) WANG C Y.
PA (ZAMB/) ZAMB I.
PI Ladd AE, Wang CY, Zamb I;

DR WPI; 94-357910/44.
PT Immunogenic luteinising hormone releasing hormone peptide(s) -
PT that suppress LHRH activity in males and females
PS Claim 8; Page 88; 213pp; English.
CC Synthetic immunogenic peptides are provided in which a universal immune
CC stimulator is linked to a peptide or protein haptens containing B cell
CC and/or cytotoxic T lymphocyte epitopes, giving a product which causes
CC potent immune responses to the coupled peptide or protein. The
CC stimulator consists of (A) a promiscuous helper T cell epitope (Th)
CC which elicits an immune response to the coupled peptide in members of
CC a heterogeneous population expressing diverse HLA phenotypes, and (B)
CC an adjuvant peptide sequence from the invasin protein of Yersinia.
CC Spacer amino acid sequences (e.g. Gly-Gly) can be provided between the
CC invasin and Th domains and between the immune stimulator and haptens
CC components. When the haptens is LHRH, then optionally the invasin domain
CC can be omitted from the immune stimulator component.
CC The present sequence represents an LHRH-containing immunogenic peptide
CC as above which can be used as a potent vaccine for treating e.g.
CC prostatic hyperplasia, androgen-dependent carcinoma, prostatic
CC carcinoma, testicular carcinoma, endometriosis, benign uterine tumours,
CC recurrent functional ovarian cysts, (severe) premenstrual syndrome or
CC oestrogen-dependent breast cancer, or for induction of infertility.
SQ Sequence 47 AA;
Query Match 100.0%; Score 104; DB 13; Length 47;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 21 qyikanskfigitel 35
QY 1 QYIKANSKFIGITEL 15
RESULT 20
ID W06131 standard; Peptide; 50 AA.
AC W06131;
DT 07-FEB-1997 (first entry)
DE Anti-cholesterol ester transfer multivalent vaccine peptide.
KW Cholesterol ester transfer protein; CETP; antigen; vaccine;
KW cardiovascular disease; atherosclerosis.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_difference 1
FT /note= "C-terminal Cys residue is present for use
FT in linking the peptide to itself or other
FT molecules"
FT 2..15
FT region /label= T-cell epitope
FT /note= "T-cell epitope comprises amino acids
FT 830-843 of tetanus toxoid protein"
FT 16..34
FT region /label= B-cell epitope
FT /note= "B-cell epitope comprises amino acids
FT 349-367 of human CETP"
FT 35..50
FT region /label= B-cell epitope
FT /note= "B-cell epitope comprises the C-terminal 16
FT amino acids of human CETP, involved in
FT neutral lipid binding or transfer activity"
PN W09534888-A1.
PD 07-NOV-1996.
PF 01-MAY-1996; U06147.
PR 01-MAY-1995; US-432483.
PA (TCEL-) T CELL SCI INC.
PI Rittershaus CW, Thomas LJ;
DR WPI; 96-506103/50.
PT Cholesterol ester transfer protein B cell epitope linked to T cell
PT epitope - used to generate vaccine to regulate CETP activity for
PT decreasing the risk of developing a cardiovascular disease e.g.
PT atherosclerosis
PS Disclosure; Page 7; 72pp; English.
CC A multivalent vaccine comprises an immunogenic helper T-cell
CC epitope of tetanus toxoid protein covalently linked to the B-cell

CC epitopes of human cholesteryl ester transfer protein (CETP) (see
CC also W06127). The vaccine elicits an immune response against
CC endogenous CETP activity, and is used to treat or prevent a
CC cardiovascular disease, such as atherosclerosis.
SQ Sequence 50 AA;

Query Match 100.0%; Score 104; DB 20; Length 50;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 qyikanskfigitel 16
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

RESULT 21
ID P70345 standard; Protein; 573 AA.
AC P70345;
DT 22-APR-1991 (first entry)
DE Portion of B fragment and all of the C fragment of tetanus toxin.
KW TT; vaccine.
OS Clostridium tetani.
PN EP-209281-A.
PD 21-JAN-1987.
PF 27-JUN-1986; 305029.
PR 28-JUN-1985; GB-016442.
PA (WELL) Wellcome Foundation Ltd.
PI Fairweather NF;
DR WPI; 87-015999/03.
DR N-PSDB; N70545.
PT Cloned DNA sequence coding for tetanus toxin - or its fragments
PT contg. epitope used to express antigens for vaccine production.
PS Claim 4; Fig 1; 36pp; English.
CC Gene product comprises a tetanus toxin fragment, which may be
CC expressed in a transformed host, and used as an antigen in vaccine
CC production, against the disease.
SQ Sequence 573 AA;

Query Match 100.0%; Score 104; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 4.95e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 88 qyikanskfigitel 102
| | | | | | | | | | | | | | |
QY 1 QYIKANSKFIGITEL 15

Search completed: Wed Aug 4 15:45:05 1999
Job time : 20 secs.

[M][E][R][E] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:48:47 1999; MasPar time 6.13 Seconds
Tabular output not generated. 169.016 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSFWLVRPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 403756 seqs, 49297578 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWU8 20:NEWU9

Statistics: Mean 22.241; Variance 74.146; scale 0.300

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Match Length DB ID	Pred. No.
1	158	100.0	21 16 US-09-171- Sequence 10, Applicati 9.62e-09
2	158	100.0	21 12 US-08-724- Sequence 5, Applicatio 9.62e-09
3	158	100.0	21 8 US-08-406- Sequence 4, Applicatio 9.62e-09
4	158	100.0	21 5 US-08-161- Sequence 66, Applicati 9.62e-09
5	158	100.0	21 8 US-08-432- Sequence 3, Applicatio 9.62e-09
6	158	100.0	21 5 US-08-161- Sequence 66, Applicati 9.62e-09
7	158	100.0	21 6 US-08-245- Sequence 14, Applicati 9.62e-09
8	158	100.0	21 8 US-08-432- Sequence 3, Applicatio 9.62e-09
9	158	100.0	21 14 US-08-945- Sequence 3, Applicatio 9.62e-09
10	158	100.0	21 3 US-07-678- Sequence 1, Applicatio 9.62e-09
11	158	100.0	21 15 US-09-049- Sequence 2, Applicatio 9.62e-09
12	158	100.0	21 15 US-09-089- Sequence 5, Applicatio 9.62e-09
13	158	100.0	22 10 US-08-577- Sequence 2, Applicatio 9.62e-09
14	158	100.0	31 5 US-08-161- Sequence 64, Applicati 9.62e-09
15	158	100.0	31 5 US-08-161- Sequence 64, Applicati 9.62e-09
16	158	100.0	32 1 PCT-US94-0 Sequence 14, Applicati 9.62e-09
17	158	100.0	32 6 US-08-229- Sequence 14, Applicati 9.62e-09
18	158	100.0	32 4 US-08-057- Sequence 14, Applicati 9.62e-09
19	158	100.0	32 9 US-08-488- Sequence 14, Applicati 9.62e-09
20	158	100.0	1315 14 US-08-913- Sequence 1, Applicatio 9.62e-09
21	152	96.2	22 6 US-08-218- Sequence 8, Applicatio 4.70e-08

22	152	96.2	22 1 PCT-US94-0 Sequence 5, Applicatio 4.70e-08
23	152	96.2	22 7 US-08-328- Sequence 8, Applicatio 4.70e-08
24	152	96.2	22 6 US-08-229- Sequence 5, Applicatio 4.70e-08
25	152	96.2	22 9 US-08-488- Sequence 8, Applicatio 4.70e-08
26	152	96.2	22 14 US-08-926- Sequence 8, Applicatio 4.70e-08
27	152	96.2	22 4 US-08-060- Sequence 8, Applicatio 4.70e-08
28	152	96.2	22 12 US-08-718- Sequence 8, Applicatio 4.70e-08
29	152	96.2	22 16 US-09-100- Sequence 41, Applicati 4.70e-08
30	152	96.2	22 4 US-08-057- Sequence 5, Applicati 4.70e-08
31	152	96.2	22 16 US-09-100- Sequence 33, Applicati 4.70e-08
32	152	96.2	32 7 US-08-328- Sequence 19, Applicati 4.70e-08
33	152	96.2	32 6 US-08-218- Sequence 19, Applicati 4.70e-08
34	152	96.2	32 12 US-08-718- Sequence 19, Applicati 4.70e-08
35	152	96.2	32 4 US-08-060- Sequence 19, Applicati 4.70e-08
36	152	96.2	34 12 US-08-718- Sequence 38, Applicati 4.70e-08
37	152	96.2	34 7 US-08-328- Sequence 38, Applicati 4.70e-08
38	132	83.5	20 14 US-08-944- Sequence 20, Applicati 8.68e-06
39	116	73.4	14 15 US-09-003- Sequence 43, Applicati 5.18e-04
40	116	73.4	14 1 PCT-US98-0 Sequence 43, Applicati 5.18e-04
41	116	73.4	14 18 US-09-321- Sequence 43, Applicati 5.18e-04
42	114	72.2	15 18 US-09-321- Sequence 44, Applicati 8.57e-04
43	114	72.2	15 15 US-09-003- Sequence 44, Applicati 8.57e-04
44	114	72.2	15 1 PCT-US98-0 Sequence 44, Applicati 8.57e-04
45	114	72.2	22 13 US-08-817- Sequence 8, Applicatio 8.57e-04

ALIGNMENTS

RESULT 1
ID US-09-171-969-10 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX

Sequence 10, Application US/09171969

Sequence 10, Application US/09171969

GENERAL INFORMATION:

APPLICANT: Thomas, Lawrence J.

TITLE OF INVENTION: PLASMID-BASED VACCINE FOR TREATING ATHEROSCLEROSIS

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 75 State Street, Suite 2300

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1807

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/171,969

FILING DATE: 01 May 1997 (01.05.97)

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/640,713

FILING DATE: 01 May 1996 (01.05.96)

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/802,967

FILING DATE: 21 February 1997 (21.02.97)

ATTORNEY/AGENT INFORMATION:

NAME: Leon R. Yankwich

REGISTRATION NUMBER: 30,237

REFERENCE/DOCKET NUMBER: TCS 414.1 PCT (05872)

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

TOPOLOGY: linear

```
CC MOLECULE TYPE: protein
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FEATURE:
CC NAME/KEY:
CC LOCATION:
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 16; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLVRVPKVSASHLE 21
   |||||
QY 1 FNNFTVSFWLVRVPKVSASHLE 21

RESULT 2
ID US-08-724-774B-5 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application US/08724774B
XX
CC Sequence 5, Application US/08724774B
CC GENERAL INFORMATION:
CC APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie,
CC APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,
CC APPLICANT: Stefan; Reed, Daryl
CC TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor
CC TITLE OF INVENTION: Rejection Antigen Precuros Mage-10,
CC TITLE OF INVENTION: Antibodies Specific To The Molecule, and
CC TITLE OF INVENTION: Uses Thereof
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Felfe & Lynch
CC STREET: 805 Third Avenue
CC CITY: New York City
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10022
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
CC COMPUTER: IBM
CC OPERATING SYSTEM: PC-DOS
CC SOFTWARE: Wordperfect
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/724,774B
CC FILING DATE: 03-October-1996
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Hanson, Norman D.
CC REGISTRATION NUMBER: 30,946
CC REFERENCE/DOCKET NUMBER: LUD 5457
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 688-9200
CC TELEFAX: (212) 838-3884
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21
CC TYPE: amino acid
CC TOPOLOGY: linear
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 12; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLVRVPKVSASHLE 21
   .|||||
QY 1 FNNFTVSFWLVRVPKVSASHLE 21
```

```
RESULT 3
ID US-08-406-916B-4 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 4, Application US/08406916B
XX
CC Sequence 4, Application US/08406916B
CC GENERAL INFORMATION:
CC APPLICANT: STEVENS, VERNON C.
CC TITLE OF INVENTION: VACCINES AND ANTIGENIC CONJUGATES
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MILLARD, SIDNEY W.
CC STREET: 7632 SLATE RIDGE BOULEVARD
CC CITY: REYNOLDSBURG
CC STATE: OHIO
CC COUNTRY: USA
CC ZIP: 43068
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB
CC COMPUTER: IBM PS/2 COMPATIBLE 486 DX 33
CC OPERATING SYSTEM: MS DOS 6.2
CC SOFTWARE: WORDPERFECT FOR MS DOS 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/406,916B
CC FILING DATE: 27 MAR 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US92/08370
CC FILING DATE: 30 SEP 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KREMBLAS, FRANCIS T., JR.
CC REGISTRATION NUMBER: 22,773
CC REFERENCE/DOCKET NUMBER: URF 2 056 3 3 3
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (614) 575 2100
CC TELEFAX: (614) 575 2149
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE:
CC DESCRIPTION: peptide
CC HYPOTHETICAL: No
CC FRAGMENT TYPE: 947-967 fragment of tetanus toxoid
CC FEATURE:
CC OTHER INFORMATION: T cell epitope
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLVRVPKVSASHLE 21
   |||||
QY 1 FNNFTVSFWLVRVPKVSASHLE 21

RESULT 4
ID US-08-161-889A-66 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 66, Application US/08161889A
XX
CC Sequence 66, Application US/08161889A
```

CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 74
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM: Floppy disk
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889A
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 66:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSEFWLRVPKVSASHLE 21
| | | | | | | | | | | | | | | | | | | | |
QY 1 FNNFTVSEFWLRVPKVSASHLE 21

RESULT 5
ID US-08-432-483-3 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 3, Application US/08432483
Sequence 3, Application US/08432483
GENERAL INFORMATION:
CC APPLICANT: Rittershaus, Charles, W.
CC TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
CC TITLE OF INVENTION: TRANSFER PROTEIN (CETP) ACTIVITY
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Banner & Allegretti, Ltd.
CC STREET: Ten South Wacker Drive
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606-7407
CC COMPUTER READABLE FORM: Floppy disk
CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 6.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/432,483
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Leon R. Vankwich
CC REGISTRATION NUMBER: 30,237
CC REFERENCE/DOCKET NUMBER: 95,179(TCS-95179)
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL:
CC ANTI-SENSE:
CC FEATURE:
CC NAME/KEY: 21-amino acid tetanus toxoid universal helper T cell
CC LOCATION:
CC PUBLICATION INFORMATION:
CC AUTHORS: Panina-Bordignon, P., et al.
CC TITLE: Universally immunogenic T cell
CC TITLE: epitopes: promiscuous binding to human MHC class II and
CC TITLE: promiscuous recognition by T cells
CC JOURNAL: European Journal of Immunology
CC VOLUME: 19
CC ISSUE:
CC PAGES: 2237-2242
CC DATE: 1989
CC RELEVANT RESIDUES IN SEQ ID NO: 3:FROM 1 TO 21
SQ SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 8; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSEFWLRVPKVSASHLE 21
| | | | | | | | | | | | | | | | | | | | |
QY 1 FNNFTVSEFWLRVPKVSASHLE 21

RESULT 6
ID US-08-161-889-66 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 66, Application US/08161889
Sequence 66, Application US/08161889
GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 66:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: linear
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 5; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
| | | | | | | | | | | | | | | | | | | | |
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 7
ID US-08-245-507-14 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX

Sequence 14, Application US/08245507
Sequence 14, Application US/08245507
GENERAL INFORMATION:
APPLICANT: Houston, Michael
APPLICANT: Zhou, Nian
APPLICANT: Kay, Cyril
APPLICANT: Hodges, Robert
APPLICANT: Cachia, Paul
APPLICANT: Irvin, Randall

TITLE OF INVENTION: Heterodimer Polypeptide Immunogen Carrier
TITLE OF INVENTION: Composition and Method
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,507
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 8900-0009
TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (415) 324-0880
CC TELEFAX: (415) 324-0960
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: T antigen, TT3 peptide
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
| | | | | | | | | | | | | | | | | | | | |
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 8
ID US-08-432-483A-3 STANDARD; PRT; 21 AA.
XX
AC xxxxxx
XX
DT
XX

Sequence 3, Application US/08432483A
Sequence 3, Application US/08432483A
GENERAL INFORMATION:
APPLICANT: Rittershaus, Charles W.
TITLE OF INVENTION: MODULATION OF CHOLESTERYL ESTER
TITLE OF INVENTION: TRANSFER PROTEIN (CETP) ACTIVITY
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: Ten South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-7407

COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/432,483A
FILING DATE: 1-May-1995
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Leon R. Yankwich
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: 95,179(TCS-95179)
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL:
ANTI-SENSE:
FEATURE:

NAME/KEY: 21-amino acid tetanus toxoid universal
NAME/KEY: helper T cell epitope.


```
CC          TELEX: 904136
CC          INFORMATION FOR SEQ ID NO: 2:
CC          SEQUENCE CHARACTERISTICS:
CC          LENGTH: 22 amino acids
CC          TYPE: amino acid
CC          STRANDEDNESS: single
CC          TOPOLOGY: linear
CC          SQ  SEQUENCE 22 AA; 2582 MW; 2920 CN;

Query Match      100.0%; Score 158; DB 10; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
   |||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 14
ID US-08-161-889-64 STANDARD; PRT; 31 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 64, Application JS/08161889
XX
CC Sequence 64, Application US/08161889
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3694 MW; 6041 CN;

Query Match      100.0%; Score 158; DB 5; Length 31;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 FNNFTVSFWLRVPKVSASHLE 27
   |||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 16
ID PCT-US94-04832A-14 STANDARD; PRT; 32 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 14, Application PC/TUS9404832A
XX
CC Sequence 14, Application PC/TUS9404832A
```

```
QY 1 FNNFTVSFWLRVPKVSASHLE 21
   |||||
RESULT 15
ID US-08-161-889A-64 STANDARD; PRT; 31 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 64, Application US/08161889A
XX
CC Sequence 64, Application US/08161889A
CC GENERAL INFORMATION:
CC APPLICANT: Geysen, H. Mario
CC APPLICANT: Rodda, Stuart J.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 74
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/161,889A
CC FILING DATE: 02-DEC-1993
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.002
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3694 MW; 6041 CN;

Query Match      100.0%; Score 158; DB 5; Length 31;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 FNNFTVSFWLRVPKVSASHLE 27
   |||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 16
ID PCT-US94-04832A-14 STANDARD; PRT; 32 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 14, Application PC/TUS9404832A
XX
CC Sequence 14, Application PC/TUS9404832A
```

CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE:
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/04832A
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME:
CC REGISTRATION NUMBER:
CC REFERENCE/DOCKET NUMBER:
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)742-4343
CC TELEFAX: (516)742-4366
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA; 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 17
ID US-08-229-275-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 14, Application US/08229275
XX
CC Sequence 14, Application US/08229275
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs as
CC TITLE OF INVENTION: vaccines for treatment of prostate cancer and induction of
CC TITLE OF INVENTION: infertility
CC NUMBER OF SEQUENCES: 53
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: M. Lisa Wilson
CC STREET: 25 Davids Drive
CC CITY: Hauppauge
CC STATE: NY
CC COUNTRY: US
CC ZIP: 11788
CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/229,275
CC FILING DATE: 13-APR-1994
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, M L
CC REGISTRATION NUMBER: 34,045
CC REFERENCE/DOCKET NUMBER: 2003Z
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)273-2828
CC TELEFAX: (516)273-1717
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA; 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 6; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.52e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 18
ID US-08-057-166-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 14, Application US/08057166
XX
CC Sequence 14, Application US/08057166
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC TITLE OF INVENTION: Immunogenic Peptides which Contain LHRH
CC TITLE OF INVENTION: and a Helper T Cell Epitope for Treatment of Prostate
CC TITLE OF INVENTION: Cancer and Induction of Infertility
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: M. Lisa Wilson, United Biomedical Inc.
CC STREET: 25 Davids Dr.
CC CITY: Hauppauge
CC STATE: NY
CC COUNTRY: U.S.A.
CC ZIP: 11788
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/057,166
CC FILING DATE: 19930427
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, M. Lisa
CC REGISTRATION NUMBER: 34,045
CC REFERENCE/DOCKET NUMBER: 2003
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (516)273-2828
CC TELEFAX: (516)273-1717
CC INFORMATION FOR SEQ ID NO: 14:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 32 AA; 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 4; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLVRPKVSASHLE 23
QY 1 FNNFTVSFWLVRPKVSASHLE 21

RESULT 19
ID US-08-488-320A-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
DT
XX
DE
XX
CC Sequence 14, Application US/08488320A
CC Sequence 14, Application US/08488320A
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic Peptides Which Contain LHRH
CC TITLE OF INVENTION: And A Helper T-Cell Epitope For Treatment Of Prostate Cancer
CC TITLE OF INVENTION: And Induction Of Infertility
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,320A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/488,351
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/446,692
CC FILING DATE: 07-JUN-1995
CC APPLICATION NUMBER: 08/229,275
CC FILING DATE: 14-APR-1994
CC APPLICATION NUMBER: 08/057,166
CC FILING DATE: 27-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C. H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146US4
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)758-4800
CC TELEFAX: (212)751-5849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide

SQ SEQUENCE 32 AA; 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 9; Length 32;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLVRPKVSASHLE 23
QY 1 FNNFTVSFWLVRPKVSASHLE 21

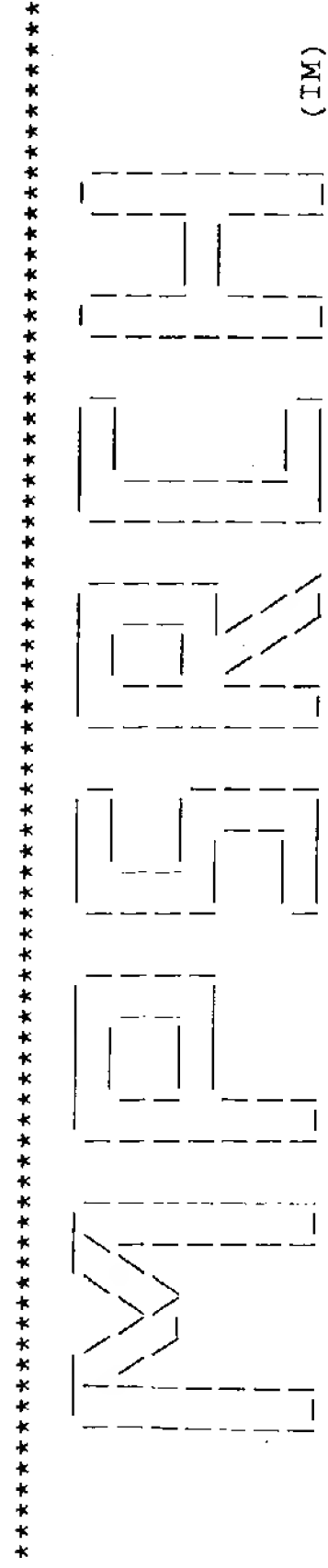
RESULT 20
ID US-08-913-880-1 STANDARD; PRT; 1315 AA.
XX
AC xxxxxx
DT
XX
DE
XX
CC Sequence 1, Application US/08913880
CC Sequence 1, Application US/08913880
CC GENERAL INFORMATION:
CC APPLICANT: MATSUDA, Morihiro
CC TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIGEN AND TETANUS
CC TITLE OF INVENTION: VACCINE
CC FILE REFERENCE: 216-380P
CC CURRENT APPLICATION NUMBER: US/08/913,880
CC CURRENT FILING DATE: 1997-09-24
CC NUMBER OF SEQ ID NOS: 9
CC SOFTWARE: PatentIn Ver. 2.0
CC SEQ ID NO 1
CC LENGTH: 1315
CC TYPE: PRT
CC ORGANISM: Clostridium tetani
CC SEQUENCE 1315 AA; 150681 MW; 9110516 CN;

Query Match 100.0%; Score 158; DB 14; Length 1315;
Best Local Similarity 100.0%; Pred. No. 9.62e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 947 FNNFTVSFWLVRPKVSASHLE 967
QY 1 FNNFTVSFWLVRPKVSASHLE 21

Search completed: Wed Aug 4 15:49:29 1999
Job time : 42 secs.

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:26:30 1999; MasPar time 2.40 Seconds
Tabular output not generated. 88.689 Million cell updates/sec

Title: >US-09-049-847-2
Description: (1-21) from US09049847.pep
Perfect Score: 158
Sequence: 1 FNNFTVSEWLRVPKVSASHLE 21

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued

1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 20.415; Variance 71.502; scale 0.286

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	158	100.0	21	2	US-08-460-	Sequence 8, Applicatio	3.25e-09
2	158	100.0	21	1	US-07-610-	Sequence 1, Applicatio	3.25e-09
3	158	100.0	21	2	US-08-661-	Sequence 12, Applicati	3.25e-09
4	158	100.0	21	3	PCT-US93-1	Sequence 66, Applicati	3.25e-09
5	158	100.0	31	3	PCT-US93-1	Sequence 64, Applicati	3.25e-09
6	158	100.0	32	2	US-08-488-	Sequence 14, Applicati	3.25e-09
7	158	100.0	32	2	US-08-446-	Sequence 14, Applicati	3.25e-09
8	158	100.0	452	1	US-07-618-	Sequence 4, Applicatio	3.25e-09
9	158	100.0	452	1	US-08-280-	Sequence 4, Applicatio	3.25e-09
10	158	100.0	452	1	US-07-618-	Sequence 2, Applicatio	3.25e-09
11	158	100.0	452	1	US-08-280-	Sequence 2, Applicatio	3.25e-09
12	158	100.0	452	1	US-08-110-	Sequence 8, Applicatio	3.25e-09
13	158	100.0	618	2	US-08-668-	Sequence 5, Applicatio	3.25e-09
14	152	96.2	22	3	PCT-US95-1	Sequence 8, Applicatio	3.25e-09
15	152	96.2	22	2	US-08-488-	Sequence 5, Applicatio	1.50e-08
16	152	96.2	22	2	US-08-446-	Sequence 5, Applicatio	1.50e-08
17	142	89.9	19	1	US-07-610-	Sequence 2, Applicatio	1.90e-07
18	135	85.4	20	2	US-08-319-	Sequence 11, Applicati	1.10e-06
19	126	79.7	17	1	US-07-610-	Sequence 3, Applicatio	1.04e-05
20	116	73.4	14	2	US-08-787-	Sequence 43, Applicati	1.23e-04
21	116	73.4	14	1	US-07-610-	Sequence 6, Applicatio	1.23e-04
22	114	72.2	15	1	US-07-610-	Sequence 4, Applicatio	2.01e-04
23	114	72.2	15	2	US-08-787-	Sequence 44, Applicati	2.01e-04

24	99	62.7	13	1	US-07-610-	Sequence 5, Applicatio	7.54e-03
25	90	57.0	438	1	US-08-480-	Sequence 23, Applicati	6.34e-02
26	90	57.0	462	1	US-08-480-	Sequence 26, Applicati	6.34e-02
27	90	57.0	1296	1	US-08-480-	Sequence 28, Applicati	6.34e-02
28	67	42.4	785	1	US-08-526-	Sequence 4, Applicatio	1.17e+01
29	67	42.4	785	2	US-08-946-	Sequence 4, Applicatio	1.17e+01
30	65	41.1	510	1	US-08-489-	Sequence 1, Applicatio	1.80e+01
31	62	39.2	297	1	US-08-176-	Sequence 16, Applicati	3.42e+01
32	62	39.2	297	2	US-08-461-	Sequence 16, Applicati	3.42e+01
33	62	39.2	1127	3	PCT-US95-0	Sequence 3, Applicatio	3.42e+01
34	59	37.3	256	3	PCT-US95-1	Sequence 67, Applicati	6.44e+01
35	58	36.7	423	2	US-08-808-	Sequence 1, Applicatio	7.94e+01
36	58	36.7	583	2	US-08-448-	Sequence 5, Applicatio	7.94e+01
37	57	36.1	104	1	US-07-942-	Sequence 16, Applicati	9.77e+01
38	57	36.1	113	2	US-08-082-	Sequence 5, Applicatio	9.77e+01
39	57	36.1	113	2	US-08-472-	Sequence 5, Applicatio	9.77e+01
40	57	36.1	113	2	US-08-477-	Sequence 5, Applicatio	9.77e+01
41	57	36.1	113	2	US-08-107-	Sequence 5, Applicatio	9.77e+01
42	57	36.1	294	1	US-08-624-	Sequence 4, Applicatio	9.77e+01
43	57	36.1	426	2	US-08-631-	Sequence 5, Applicatio	9.77e+01
44	56	35.4	256	3	PCT-US95-0	Sequence 3, Applicatio	1.20e+02
45	56	35.4	2910	3	PCT-US95-0	Sequence 157, Applicat	1.20e+02

ALIGNMENTS

RESULT	1				
ID	US-08-460-502-8	STANDARD;	PRT;	21	AA.
XX					
AC	xxxxxxx				
XX					
DT					
XX					

Sequence 8, Application US/08460502

Sequence 8, Application US/08460502

Patent No. 5843464

GENERAL INFORMATION:

APPLICANT: Bakaletz, Lauren O.

APPLICANT: Kaumaya, Parvin T.

TITLE OF INVENTION: Synthetic Chimeric Fimbrin Peptides

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Calfee, Halter and Griswold

STREET: 800 Superior Avenue

CITY: Cleveland

STATE: Ohio

COUNTRY: U.S.A.

ZIP: 44114-2688

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,502

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Golrick, Mary E.

REGISTRATION NUMBER: 34,829

REFERENCE/DOCKET NUMBER: 22727/00120

TELEPHONE: (216) 622-8458

TELEFAX: (216) 241-0816

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 21 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
|||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 2
ID US-07-610-525-1 STANDARD; PRT; 21 AA.

AC xxxxxx

Sequence 1, Application US/07610525

Sequence 1, Application US/07610525
Patent No. 5196512

GENERAL INFORMATION:
APPLICANT: BIANCHI Elisabetta
APPLICANT: PESSI Antonello
APPLICANT: CORRADIN Giampietro
TITLE OF INVENTION: SYNTHETIC PEPTIDES USEFUL AS
TITLE OF INVENTION: UNIVERSAL CARRIERS FOR THE PREPARATION OF IMMUNOGENIC CONJ
TITLE OF INVENTION: AND THEIR USE IN THE DEVELOPMENT OF SYNTHETIC VACCINES.
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:
ADDRESSEE: SHEA & GOULD
STREET: 1251 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES
ZIP: 10020-1193

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/610,525
FILING DATE: 19901108
CLASSIFICATION: 424
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-827-3000
TELEFAX: 212-840-6702
TELEX: 423973

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acid residues
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: no
FRAGMENT TYPE: internal fragment
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
|||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 3
ID US-08-661-052-12 STANDARD; PRT; 21 AA.

XX xxxxxx

AC xxxxxx

XX
DT
XX
DE
XX
Sequence 12, Application US/08661052

Sequence 12, Application US/08661052
Patent No. 5837243

GENERAL INFORMATION:
APPLICANT: Yashwant M. Deo
APPLICANT: Joel Goldstein
APPLICANT: Robert Graziano
APPLICANT: Chezian Somasundaram
TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
TITLE OF INVENTION: OF ANTI-FC RECEPTOR ANTIBODIES
NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/661,052

FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/484,172
FILING DATE: 07-JUNE-1995

ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: MXI-043CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
SEQUENCE 21 AA; 2479 MW; 2810 CN;

Query Match 100.0%; Score 158; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSFWLRVPKVSASHLE 21
|||||
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 4
ID PCT-US93-11703-66 STANDARD; PRT; 21 AA.

XX xxxxxx

Sequence 66, Application PC/TUS9311703

Sequence 66, Application PC/TUS9311703
GENERAL INFORMATION:

APPLICANT: Chiron Mimotopes Pty. Ltd.
TITLE OF INVENTION: T-Cell Epitopes
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCI/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 66:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 21 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 21 AA; 2479 MW; 2810 CN;
SQ

Query Match 100.0%; Score 158; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 FNNFTVSWLVRVPKVSASHLE 21
|||||
QY 1 FNNFTVSWLVRVPKVSASHLE 21

RESULT 5
ID PCT-US93-11703-64 STANDARD; PRT; 31 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 64, Application PC/TUS9311703
XX
CC Sequence 64, Application PC/TUS9311703
CC GENERAL INFORMATION:
CC APPLICANT: Chiron Mimotopes Pty. Ltd.
CC TITLE OF INVENTION: T-Cell Epitopes
CC NUMBER OF SEQUENCES: 75
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Grant D. Green
CC STREET: 4560 Horton St.
CC CITY: Emeryville
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94608
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30B
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/11703
CC FILING DATE: 28-DEC-1993
CC CLASSIFICATION:

CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/984,852
CC FILING DATE: 02-DEC-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Green, Grant D.
CC REGISTRATION NUMBER: 31,259
CC REFERENCE/DOCKET NUMBER: 0222.101
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 510-601-2706
CC TELEFAX: 510-655-3542
CC INFORMATION FOR SEQ ID NO: 64:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 31 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 31 AA; 3694 MW; 6041 CN;
SQ

Query Match 100.0%; Score 158; DB 3; Length 31;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 7 FNNFTVSWLVRVPKVSASHLE 27
|||||
QY 1 FNNFTVSWLVRVPKVSASHLE 21

RESULT 6
ID US-08-488-351A-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 14, Application US/08488351A
XX
CC Sequence 14, Application US/08488351A
CC Patent No. 5843446
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/488,351A
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/229,275
CC FILING DATE: 14-APR-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992
CC


```
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA; 3789 MW; 6283 CN;

Query Match 100.0%; Score 158; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21
|||||

RESULT 7
ID US-08-446-692-14 STANDARD; PRT; 32 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 14, Application US/08446692
XX
CC Sequence 14, Application US/08446692
CC Patent No. 5759551
CC GENERAL INFORMATION:
CC APPLICANT: Ladd, Anna
CC APPLICANT: Wang, Chang Yi
CC APPLICANT: Zamb, Timothy
CC TITLE OF INVENTION: Immunogenic LHRH peptide constructs
CC TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
CC NUMBER OF SEQUENCES: 114
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Maria C.H. Lin
CC STREET: 345 Park Avenue
CC CITY: New York
CC STATE: NY
CC COUNTRY: US
CC ZIP: 10154-0053
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/446,692
CC FILING DATE: 7-JUN-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 32 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 32 AA; 3789 MW; 6283 CN;
```

```
Query Match 100.0%; Score 158; DB 2; Length 32;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSFWLRVPKVSASHLE 23
QY 1 FNNFTVSFWLRVPKVSASHLE 21
|||||

RESULT 8
ID US-07-618-312A-4 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 4, Application US/07618312A
XX
CC Sequence 4, Application US/07618312A
CC Patent No. 5389540
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: 14th Floor
CC STREET: 2200 Clarendon Boulevard,
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/618,312A
CC FILING DATE: 19910516
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Crawford Mr, Arthur R
CC REGISTRATION NUMBER: 25,327
CC REFERENCE/DOCKET NUMBER: 510-51
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 0101 703 8750400
CC TELEFAX: 0101 703 5253468
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51770 MW; 1065961 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21
|||||
```

```

XX
AC
XX
XX
DT
XX
DE
XX
Sequence 2, Application US/07618312A
Sequence 2, Application US/07618312A
Patent No. 5389540
GENERAL INFORMATION:
APPLICANT: Makoff Dr, Andrew J
APPLICANT: Romanos Dr, Michael A
APPLICANT: Clare Dr, Jeffrey J
APPLICANT: Fairweather Dr, Neil F
TITLE OF INVENTION: VACCINES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: 14th Floor
STREET: 2200 Clarendon Boulevard,
CITY: Arlington,
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/618,312A
FILING DATE: 19910516
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 8926832.0
FILING DATE: 28-NOV-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9006097.1
FILING DATE: 17-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Crawford Mr, Arthur R
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 510-51
TELECOMMUNICATION INFORMATION:
TELEPHONE: 0101 703 8750400
TELEFAX: 0101 703 5253458
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
SQ SEQUENCE 452 AA; 51797 MW; 1065526 CN;
Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 84 FNNFTVSFWLVRVPKVSASHLE 104
|||||
QY 1 FNNFTVSFWLVRVPKVSASHLE 21
RESULT 11
ID US-08-280-228-2 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 2, Application US/08280228
DE
XX
Sequence 2, Application US/08280228

```

```
CC Patent No. 5571694
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51797 MW; 1065526 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 12
ID US-08-110-786A-8 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 8, Application US/08110786A
XX
CC Sequence 8, Application US/08110786A
CC Patent No. 5443966
CC GENERAL INFORMATION:
CC APPLICANT: FAIRWEATHER, Neil Fraser
CC APPLICANT: MAKOFF, Andrew Joseph
```

```
CC TITLE OF INVENTION: Expression of tetanus toxin fragment C
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Nixon & Vanderhye P.C.
CC STREET: 1100 No. 5443966th Glebe Road
CC CITY: Arlington
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/110,786A
CC FILING DATE: 23-AUG-1993 1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/777,337
CC FILING DATE: 29-NOV-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/GB90/00943
CC FILING DATE: 20-JUN-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8914122.0
CC FILING DATE: 20 June 1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mary J. Wilson
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-134
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51784 MW; 1060967 CN;

Query Match 100.0%; Score 158; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 84 FNNFTVSFWLRVPKVSASHLE 104
QY 1 FNNFTVSFWLRVPKVSASHLE 21

RESULT 13
ID US-08-668-381A-5 STANDARD; PRT; 618 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application US/08668381A
XX
CC Sequence 5, Application US/08668381A
CC Patent No. 5780024
CC GENERAL INFORMATION:
CC APPLICANT: Brown, Robert H.
CC APPLICANT: Fishman, Paul S.
CC APPLICANT: Francis, Jonathan W.
CC APPLICANT: Hosler, Betsy A.
CC TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
CC TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
```

CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/C00,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul I.
CC REGISTRATION NUMBER: 30,164
CC REFERENCE/DOCKET NUMBER: 00786/269001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 618 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 618 AA; 68895 MW; 1991829 CN;

Query Match 100.0%; Score 158; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 3.25e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 250 FNNFTVSFVLRVPKVSASHLE 270
QY 1 FNNFTVSFVLRVPKVSASHLE 21

RESULT 14
ID PCT-US95-13841-8 STANDARD; PRT; 22 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 8, Application PC/TUS9513841
Sequence 8, Application PC/TUS9513841
GENERAL INFORMATION:
APPLICANT: United Biomedical Inc; Walfield, Alan M.;
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Synthetic IgE Membrane Anchor
TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13841
FILING DATE: 25-OCT-1995

CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/328,519
CC FILING DATE: 25-OCT-1994
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lin, Maria C.H.
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4117
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-758-4800
CC TELEFAX: 212-751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 22 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 22 AA; 2606 MW; 3203 CN;

Query Match 96.2%; Score 152; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.50e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFIVSFVLRVPKVSASHL 22
QY 1 FNNFIVSFVLRVPKVSASHL 20

RESULT 15
ID US-08-488-351A-5 STANDARD; PRT; 22 AA.
XX xxxxxx

Sequence 5, Application US/08488351A
Sequence 5, Application US/08488351A
Patent No. 5843446

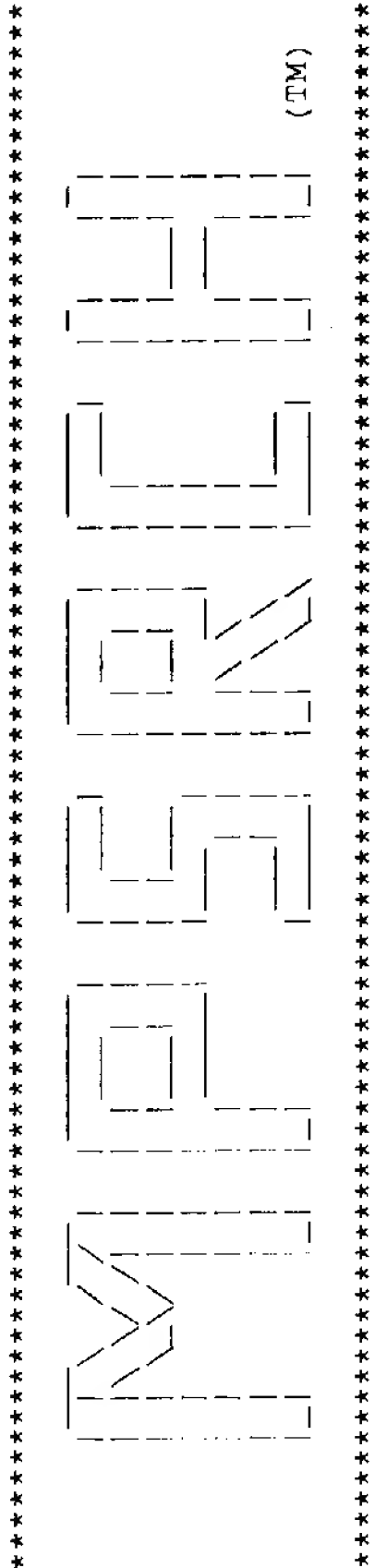
GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic LHRH peptide constructs
TITLE OF INVENTION: and synthetic universal immune stimulants for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,351A
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,275
FILING DATE: 14-APR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/057,166
CC FILING DATE: 27-APR-1992
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Maria C.H. Lin
CC REGISTRATION NUMBER: 29,323
CC REFERENCE/DOCKET NUMBER: 1151-4146 US2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212)415-8745
CC TELEFAX: (516)751-6849
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 22 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 22 AA; 2606 MW; 3203 CN;

Query Match 96.2%; Score 152; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.50e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 FNNFTVSEWLRVPKVSASHL 22
QY 1 FNNFTVSEWLRVPKVSASHL 20

Search completed: Wed Aug 4 15:26:35 1999
Job time : 5 secs.



24	48	58.5	296	2	S75858	glutamine-binding pro	1.88e+01
25	48	58.5	509	1	DENTN4	NADH dehydrogenase (u	1.88e+01
26	48	58.5	802	1	A26343	ADE5 multifunctional	1.88e+01
27	47	57.3	84	1	QIBPS2	uracil-DNA glycosylas	2.97e+01
28	47	57.3	299	2	B70951	hypothetical protein	2.97e+01
29	47	57.3	378	2	B38178	tela protein - plasm	2.97e+01
30	47	57.3	514	2	A35658	transcription factor	2.97e+01
31	47	57.3	562	2	H70688	hypothetical protein	2.97e+01
32	47	57.3	680	2	A28121	major merozoite surfa	2.97e+01
33	47	57.3	796	1	JV0107	glucose dehydrogenase	2.97e+01
34	47	57.3	821	1	A39616	protein kinase RAD53	2.97e+01
35	47	57.3	1379	2	S01254	hepatocyte growth fac	2.97e+01
36	47	57.3	1772	2	A45532	major merozoite surfa	2.97e+01
37	46	56.1	250	2	D69182	conserved hypotheticala	4.64e+01
38	46	56.1	282	2	C69553	conserved hypotheticala	4.64e+01
39	46	56.1	288	1	GNBPI4	gene 9 protein - phag	4.64e+01
40	46	56.1	320	2	D71098	hypothetical protein	4.64e+01
41	46	56.1	397	2	I64122	tryptophan synthase (4.64e+01
42	46	56.1	655	1	ALKBG	cyclomaltodextrin glu	4.64e+01
43	46	56.1	791	1	CGB01S	collagen alpha 1(I) c	4.64e+01
44	46	56.1	1184	1	A34795	kinesin-related prote	4.64e+01
45	46	56.1	2180	2	A47651	zinc-finger protein -	4.64e+01

ALIGNMENTS

RESULT	1	BTCLTN	#type complete
ENTRY		tentoxylysin (EC 3.4.24.68)	precursor - Clostridium tetani
TITLE		tetanus neurotoxin	
ALITERNAIE_NAMES		#formal_name Clostridium tetani	
ORGANISM		31-Mar-1988 #sequence_revision 31-Mar-1988	#text_change
DATE		26-Feb-1999	
ACCESSIONS		A25689; A25757; A25194; B25194; A60759; S69348; S09364	
REFERENCE		A25689	
#authors		Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.; Habermann, E.; Niemann, H.	
#journal		EMBO J. (1986) 5:2495-2502	
#title		Tetanus toxin: primary structure, expression in E. coli, and homology with botulinum toxins.	
#cross-references		MUID:87053814	
#accession		A25689	
#molecule_type		DNA	
#residues		1-1315	#label EIS
#cross-references		GB:X04436; NID:g40769; PID:g40770	
REFERENCE		A25757	
#authors		Fairweather, N.F.; Lyness, V.A.	
#journal		Nucleic Acids Res. (1986) 14:7809-7812	
#title		The complete nucleotide sequence of tetanus toxin.	
#cross-references		MUID:87040747	
#accession		A25757	
#molecule_type		DNA	
#residues		1-1315	#label FAI
#cross-references		GB:X06214; NID:g40773; PID:g40774	
#experimental_source		strain CN3911	
REFERENCE		A25194	
#authors		Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.	
#journal		J. Bacteriol. (1986) 165:21-27	
#title		Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C in Escherichia coli.	
#cross-references		MUID:86085672	
#accession		A25194	
#molecule_type		DNA	
#residues		743-1315	#label FA2
#cross-references		GB:M12739; NID:g144920; PID:g144921	
#accession		B25194	
#molecule_type		protein	
#residues		865-894	#label FA3
REFERENCE		A60759	
#authors		Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.	
#journal		Infect. Immun. (1989) 57:3588-3593	
#title		Isolation, purification, and characterization of fragment B,	

```
#cross-references MUID:90035436
#accession A60759
##molecule_type protein
##residues 461-475 ##label MAT
REFERENCE JS0098
#authors Demotz, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.;
Widmann, C.; Corradin, G.
#journal J. Immunol. (1989) 142:394-402
#title Delineation of several DR-restricted tetanus toxin T cell
epitopes.
#cross-references MUID:89093918
#contents annotation; epitope region
REFERENCE S27125
#authors Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de
Laureto, P.P.; DasGupta, B.R.; Montecucco, C.
#journal Nature (1992) 359:832-835
#title Tetanus and botulinum-B neurotoxins block neurotransmitter
release by proteolytic cleavage of synaptobrevin.
#cross-references MUID:93063293
#contents annotation
REFERENCE S69348
#authors de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.;
Montecucco, C.
#journal Eur. J. Biochem. (1995) 229:61-69
#title Structural studies on the zinc-endopeptidase light chain of
tetanus neurotoxin.
#cross-references MUID:95262688
#accession S69348
##molecule_type protein
##residues 2-31 ##label DEF
COMMENT The source of this protein was an extrachromosomal plasmid.
COMMENT The precursor is cleaved by endogenous proteinase activity to form
light (fragment A) and heavy (fragment B.C) chains that are
covalently linked by an interchain disulfide bond (the individual
chains are not toxic when separated). The amino end of the heavy
chain (fragment B) can be separated from the carboxyl end
(fragment C) by papain.
COMMENT Fragment B forms ion channels in a lipid bilayer. Fragment C binds
to gangliosides and may target the toxin to the motor end plate.
COMMENT Fragment A is a zinc-dependent endopeptidase.
COMMENT This potent neurotoxin binds to peripheral neuronal synapses, is
internalized, and moves by retrograde transport up the axon into
the spinal cord where it can move between postsynaptic and
presynaptic neurons. It inhibits neurotransmitter release by
proteolytic cleavage of synaptobrevin (vesicle-associated
membrane protein 2).
FUNCTION
#description blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide
bond in synaptobrevin 2
CLASSIFICATION #superfamily tetanus toxin
KEYWORDS hydrolase; metalloproteinase; neurotoxin; transmembrane
protein; zinc
FEATURE
2-457
#product tentoxylisin light chain (fragment A) #status
predicted #label TTL\
461-1315 #product tentoxylisin heavy chain (fragment B.C) #status
experimental #label TTH\
461-864 #domain channel forming (fragment B) #status predicted
#label TXB\
865-1315 #domain ganglioside binding (fragment C) #status
predicted #label TXC\
233,237 #binding_site zinc (His) #status predicted\
234 #active_site Glu #status predicted
SUMMARY #length 1315 #molecular-weight 150681 #checksum 4853
Query Match 100.0%; Score 82; DB 1; Length 1315;
Best Local Similarity 100.0%; Pred. No. 1.58e-07;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1273 GQIGNDPNRDIL 1284
||| ||||| |||
QY 1 GQIGNDPNRDIL 12
```

```
RESULT 2
ENTRY JN0611 #type complete
TITLE probable transcription factor DdTEP2 - slime mold
ALTERNATE_NAMES (Dictyostelium discoideum)
ORGANISM Tat-binding protein 2
#formal_name Dictyostelium discoideum
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
04-Sep-1998
ACCESSIONS JN0611
REFERENCE JN0610
#authors Shaw, D.R.; Ennis, H.L.
#journal Biochem. Biophys. Res. Commun. (1993) 193:1291-1296
#title Molecular cloning and developmental regulation of
Dictyostelium discoideum homologues of the human and yeast
HIV1 Tat-binding protein.
#cross-references MUID:93312322
#accession JN0611
##status nucleic acid sequence not shown
##molecule_type mRNA
##residues 1-403 ##label SHA
##cross-references GB:L16578; NID:g290054; PID:g290055
CLASSIFICATION #superfamily ATP-dependent 26S proteinase;
FtsH/SEC18/CDC48-type ATP-binding domain homology
KEYWORDS ATP; nucleus; P-loop
FEATURE
164-374 #domain FtsH/SEC18/CDC48-type ATP-binding domain
homology #label VATP\
191-198 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 403 #molecular-weight 45542 #checksum 8766
Query Match 63.4%; Score 52; DB 2; Length 403;
Best Local Similarity 63.6%; Pred. No. 2.85e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 178 QIGIDPPRGVL 188
||| ||| |||
QY 2 QIGNDPNRDIL 12
RESULT 3
ENTRY S69678 #type complete
TITLE 26S proteasome regulatory particle chain RPT3 - yeast
(Saccharomyces cerevisiae)
ORGANISM protein YDR394w; YTA2 protein
#formal_name Saccharomyces cerevisiae
DATE 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change
13-Sep-1998
ACCESSIONS S69678; S46606; S34353
REFERENCE S69655
#authors Dietrich, F.S.
#submission submitted to the EMBL Data Library, July 1995
#description The sequence of S. cerevisiae cosmids 9481, 9509, 9926, 9461,
and lambda 3641.
#accession S69678
##molecule_type DNA
##residues 1-428 ##label DIE
##cross-references EMBL:U32274; NID:g927313; PID:g927327; MIPS:YDR394w
REFERENCE S46605
#authors Schnall, R.; Mannhaupt, G.; Stucka, R.; Tauer, R.; Ehnlé, S.;
Schwarzlose, C.; Vetter, I.; Feldmann, H.
#journal Yeast (1994) 10:1141-1155
#title Identification of a set of yeast genes coding for a novel
family of putative ATPases with high similarity to
constituents of the 26S protease complex.
#cross-references MUID:95274317
#accession S46606
##status nucleic acid sequence not shown
##molecule_type DNA
##residues 1-341,'Y',343-428 ##label SCH
##cross-references EMBL:X73570; NID:g313879; PID:g313880
GENETICS
```

```
#gene SGD:RPT3; YTA2; YNT1
#cross-references SGD:S0002802; MIPS:YDR394w
#map_position 4R
CLASSIFICATION #superfamily ATP-dependent 26S proteinase;
                  FtsH/SEC18/CDC48-type ATP-binding domain homology
                  nucleus; P-loop
KEYWORDS
FEATURE
186-397 #domain FtsH/SEC18/CDC48-type ATP-binding domain
        homology #label VATP\
213-220 #region nucleotide-binding motif A (P-loop)
SUMMARY #length 428 #molecular-weight 47893 #checksum 5711

Query Match 63.4%; Score 52; DB 2; Length 428;
Best Local Similarity 63.6%; Pred. No. 2.85e+00;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 200 QIGIDPPRGVL 210
   ||||| | :|
QY 2 QIGNDPNRDIL 12

RESULT 4
ENTRY GNVVY #type complete
TITLE genome polyprotein - yellow fever virus (strain 17D)
CONTAINS capsid protein C; envelope protein M; major envelope protein
          E; nonstructural protein NS1; nonstructural protein NS2a;
          nonstructural protein NS2b; nonstructural protein NS3;
          nonstructural protein NS4a; nonstructural protein NS4b;
          nonstructural protein NS5
ORGANISM #formal_name yellow fever virus
DATE 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change
        15-May-1998
ACCESSIONS A03914
REFERENCE A03914
#authors Rice, C.M.; Lencches, E.M.; Eddy, S.R.; Shin, S.J.; Sheets,
          R.L.; Strauss, J.H.
#journal Science (1985) 229:726-733
#title Nucleotide sequence of yellow fever virus: implications for
        flavivirus gene expression and evolution.
#cross-references MUID:85272570
#accession A03914
#molecule_type genomic RNA
#residues 1-3411 #label RIC
CLASSIFICATION #superfamily yellow fever virus genome polyprotein; DEAD/H
                box helicase homology
KEYWORDS ATP; capsid protein; envelope protein; glycoprotein;
          nonstructural protein; P-loop; polyprotein; transmembrane
          protein
FEATURE
2-210 #product capsid protein C #status predicted #label CPC\
211-285 #product envelope protein M #status predicted #label
        EPM\
249-269 #domain transmembrane #status predicted #label TM1\
271-285 #domain transmembrane #status predicted #label TM2\
286-778 #product major envelope protein E #status predicted
        #label MEE\
740-753 #domain transmembrane #status predicted #label TM3\
755-778 #domain transmembrane #status predicted #label TM4\
779-1187 #product nonstructural protein NS1 #status predicted
        #label NS1\
1159-1180 #domain transmembrane #status predicted #label TM5\
1188-1354 #product nonstructural protein NS2a #status predicted
        #label N2A\
1355-1484 #product nonstructural protein NS2b #status predicted
        #label N2B\
1485-2107 #product nonstructural protein NS3 #status predicted
        #label NS3\
1682-1951 #domain DEAD/H box helicase homology #label DEAD\
1682-1689 #region nucleotide-binding motif A (P-loop)\
1769-1774 #region nucleotide-binding motif B\
1773-1776 #region DEAH motif\
2108-2394 #product nonstructural protein NS4a #status predicted
        #label N4A\
```

```
2395-2506 #product nonstructural protein NS4b #status predicted
        #label N4B\
2507-3411 #product nonstructural protein NS5 #status predicted
        #label NS5\
134,150,172,266,
594,755,908,986,
1796,2062,2320,
2346,2408,2467,
2720,2734,2740 #binding_site carbohydrate (Asn) (covalent) #status
        predicted
SUMMARY #length 3411 #molecular-weight 379516 #checksum 8719

Query Match 63.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 2.85e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGRNPNRD 1956
   ||| :|||
QY 1 QIGNDPNRD 10

RESULT 5
ENTRY GNVVYP #type complete
TITLE genome polyprotein - yellow fever virus (strain Pasteur
        17D-204)
CONTAINS capsid protein C; envelope protein M; major envelope protein
          E; nonstructural protein NS1; nonstructural protein NS2a;
          nonstructural protein NS2b; nonstructural protein NS3;
          nonstructural protein NS4a; nonstructural protein NS4b;
          nonstructural protein NS5
ORGANISM #formal_name yellow fever virus
DATE 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change
        02-Jul-1998
ACCESSIONS S07757
REFERENCE S07757
#authors Dupuy, A.; Despres, P.; Cahour, A.; Girard, M.; Bouloy, M.
#journal Nucleic Acids Res. (1989) 17:3989
#title Nucleotide sequence comparison of the genome of two 17D-204
        yellow fever vaccines.
#cross-references MUID:89282413
#accession S07757
#status nucleic acid sequence not shown; translation not shown
#molecule_type genomic RNA
#residues 1-3411 #label DUP
#cross-references EMBL:X15062
#note the nucleotide sequence was submitted to the EMBL Data
        Library, April 1989, in computer-readable form
CLASSIFICATION #superfamily yellow fever virus genome polyprotein; DEAD/H
                box helicase homology
KEYWORDS ATP; capsid protein; envelope protein; glycoprotein;
          nonstructural protein; P-loop; polyprotein; transmembrane
          protein
FEATURE
2-210 #product capsid protein C #status predicted #label CPC\
105-125 #domain transmembrane #status predicted #label TM1\
211-285 #product envelope protein M #status predicted #label
        EPM\
271-289 #domain transmembrane #status predicted #label TM2\
286-778 #product major envelope protein E #status predicted
        #label MEE\
736-753 #domain transmembrane #status predicted #label TM3\
756-778 #domain transmembrane #status predicted #label TM4\
779-1187 #product nonstructural protein NS1 #status predicted
        #label NS1\
1133-1151 #domain transmembrane #status predicted #label TM5\
1160-1179 #domain transmembrane #status predicted #label TM6\
1188-1354 #product nonstructural protein NS2a #status predicted
        #label N2A\
1355-1484 #product nonstructural protein NS2b #status predicted
        #label N2B\
1485-2107 #product nonstructural protein NS3 #status predicted
        #label NS3\
1682-1951 #domain DEAD/H box helicase homology #label DEAD\
```

```
1682-1689      #region nucleotide-binding motif A (P-loop)\
1769-1774      #region nucleotide-binding motif B\
1773-1776      #region DEAH motif\
2108-2394      #product nonstructural protein NS4a #status predicted
                #label N4A\
2395-2506      #product nonstructural protein NS4b #status predicted
                #label N4B\
2507-3411      #product nonstructural protein NS5 #status predicted
                #label NS5\

134,150,172,594,
908,986,1796,2062,
2320,2346,2408,
2467,2720,2734,
2740

SUMMARY      #length 3411 #molecular-weight 379528 #checksum 8417
Query Match      63.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 2.85e+00;
Matches      7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGNPNRND 1956
|:|:|:|:|
QY 1 GQIGNDPNRD 10

RESULT 6      #type complete
ENTRY      GTP cyclohydrolase I (f0LE) RP383 - Rickettsia prowazekii
TITLE      #formal_name Rickettsia prowazekii
ORGANISM    21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
DATE        21-Nov-1998
ACCESSIONS  F71695
REFERENCE    A71630
#authors    Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
            Sicheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
            Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
            C.G.
#journal    Nature (1998) 396:133-140
#title      The genome sequence of Rickettsia prowazekii and the origin
            of mitochondria.
#accession  F71695
#status     preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues   1-190 #label AND
#cross-references GB:AJ235271; GB:AJ235269; NID:g3860788; PID:e1342684;
            PID:g3860940
#experimental_source strain Madrid E

GENETICS
#gene       f0LE; RP383
SUMMARY     #length 190 #molecular-weight 21790 #checksum 3341
Query Match      62.2%; Score 51; DB 2; Length 190;
Best Local Similarity 70.0%; Pred. No. 4.61e+00;
Matches      7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 IGEDPNRGL 29
|:|:|:|:|
QY 3 IGNDPNRDIL 12

RESULT 7
ENTRY      A38256      #type complete
TITLE      GTP cyclohydrolase I (BC 3.5.4.16) mtrA - Bacillus subtilis
ORGANISM    #formal_name Bacillus subtilis
DATE        14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change
            24-Sep-1998
ACCESSIONS  A38256; A39409; G69661
REFERENCE    A38256
#authors    Gollnick, P.; Ishino, S.; Kuroda, M.I.; Henner, D.J.;
            Yancofsky, C.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1990) 87:8726-8730
```

```
#title      The mtr locus is a two-gene operon required for transcription
            attenuation in the trp operon of Bacillus subtilis.
#cross-references MUID:91062353
#accession    A38256
            ##molecule_type DNA
            ##residues 1-190 #label GOL
            ##cross-references GB:M37320; NID:g143230; PID:g143231
REFERENCE    A39409
#authors    Micka, B.; Groch, N.; Heinemann, U.; Marahiel, M.A.
#journal    J. Bacteriol. (1991) 173:3191-3198
#title      Molecular cloning, nucleotide sequence, and characterization
            of the Bacillus subtilis gene encoding the DNA-binding
            protein HBSu.
#cross-references MUID:91216992
#accession    A39409
            ##molecule_type DNA
            ##residues 1-19 #label MIC
            ##cross-references GB:X52418
REFERENCE    A69580
#authors    Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignelli, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
            S.Y.; Glaser, P.; Goffeau, A.; Gollightly, E.J.; Grandi, G.;
            Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
            A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
            Yoshikawa, H.; Danchin, A.
            Nature (1997) 390:249-256
            The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
#cross-references MUID:98044033
#accession    G69661
            ##status     nucleic acid sequence not shown; translation not shown
            ##molecule_type DNA
            ##residues 1-190 #label KUN
            ##cross-references GB:Z99115; GB:AL009126; NID:g2634478; PID:e1183723;
            PID:g2634696
            ##experimental_source strain 168
COMMENT      This enzyme catalyzes the first step of the synthesis of
            tetrahydrobiopterin, a cofactor in the synthesis of aromatic
            amino acids.
GENETICS
#gene         mtrA
#note         the two genes in this methyltryptophan resistance (mtr)
            operon are mtrA and mtrB
CLASSIFICATION #superfamily GIP cyclohydrolase I
```


KEYWORDS hydrolase
SUMMARY #length 190 #molecular-weight 21219 #checksum 1098

Query Match 52.2%; Score 51; DB 2; Length 190;
Best Local Similarity 70.0%; Pred. No. 4.61e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 IGEDPNREGL 29
||:||||:|
QY 3 IGNDPNRDIL 12

RESULT 8
ENTRY MFNZNC #type complete
TITLE matrix protein - Newcastle disease virus (strain Beaudette C)
ORGANISM #formal_name Newcastle disease virus
DATE 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 20-Mar-1998

ACCESSIONS A26111
REFERENCE A26111
#authors Chambers, P.; Millar, N.S.; Platt, S.G.; Emmerson, P.T.
#journal Nucleic Acids Res. (1986) 14:9051-9061
#title Nucleotide sequence of the gene encoding the matrix protein of Newcastle disease virus.
#cross-references MUID:87066775
#accession A26111
##molecule_type genomic RNA
##residues 1-364 ##label CHA
##cross-references GB:X04687; NID:g60940; PID:g60941

GENETICS
#gene M
CLASSIFICATION #superfamily parainfluenza virus matrix protein
KEYWORDS matrix protein
SUMMARY #length 364 #molecular-weight 39604 #checksum 4208

Query Match 52.2%; Score 51; DB 1; Length 364;
Best Local Similarity 41.7%; Pred. No. 4.61e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
|:|:|:|:|
QY 1 QGIGNDPNRDIL 12

RESULT 9
ENTRY A39094 #type complete
TITLE glucan endo-1,3-beta-glucosidase (EC 3.2.1.1-) precursor - Oerskovia xanthineolytica
ORGANISM #formal_name Oerskovia xanthineolytica
DATE 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 31-Oct-1997

ACCESSIONS A39094
REFERENCE A39094
#authors Shen, S.H.; Chretien, P.; Bastien, L.; Slilaty, S.N.
#journal J. Biol. Chem. (1991) 266:1058-1063
#title Primary sequence of the glucanase gene from Oerskovia xanthineolytica. Expression and purification of the enzyme from Escherichia coli.
#cross-references MUID:91093212
#accession A39094
##status preliminary
##molecule_type DNA
##residues 1-548 ##label SHE
##cross-references GB:M60826; GB:M38734; NID:g150444; PID:g150445
KEYWORDS glycosidase; hydrolase
SUMMARY #length 548 #molecular-weight 58088 #checksum 8536

Query Match 62.2%; Score 51; DB 2; Length 548;
Best Local Similarity 63.5%; Pred. No. 4.61e+00;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 138 QNDSDPNRNIL 148
|:|:|:|:|

QY 2 QIGNDPNRDIL 12

RESULT 10
ENTRY S64101 #type complete
TITLE PAN2 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein G3165; protein YGL094c
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 17-May-1996 #sequence_revision 17-May-1996 #text_change 06-Feb-1998

ACCESSIONS S64101
REFERENCE S64071
#authors Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
#submission submitted to the Protein Sequence Database, May 1996
#accession S64101
##molecule_type DNA
##residues 1-1115 ##label RIE
##cross-references EMBL:Z72616; NID:g1322626; PID:e243300; PID:g1322627; MIPS:YGL094c
##experimental_source strain S288C

GENETICS
#gene SGD:PAN2
#cross-references SGD:S0003062; MIPS:YGL094c
#map_position 7L
SUMMARY #length 1115 #molecular-weight 127038 #checksum 4836

Query Match 62.2%; Score 51; DB 2; Length 1115;
Best Local Similarity 50.0%; Pred. No. 4.61e+00;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 68 GHIGNSVXDIL 79
|:|:|:|:|
QY 1 QGIGNDPNRDIL 12

RESULT 11
ENTRY JQ1895 #type complete
TITLE genome polyprotein - turnip mosaic virus (strain Quebec)
CONTAINS coat protein; cytoplasmic inclusion protein; helper component protein; nuclear inclusion a protein; nuclear inclusion b protein; P1 protein; P3 protein; p6K1 protein; p6K2 protein; VPg protein
ORGANISM #formal_name turnip mosaic virus, TuMV
DATE 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 05-Jun-1998

ACCESSIONS JQ1895
REFERENCE JQ1895
#authors Nicolas, O.; Laliberte, J.F.
#journal J. Gen. Virol. (1992) 73:2785-2793
#title The complete nucleotide sequence of turnip mosaic potyvirus RNA.
#accession JQ1895
##molecule_type mRNA
##residues 1-3163 ##label NIC
##cross-references DDBJ:D10927; NID:g222660; PID:d1002200; PID:g222661
CLASSIFICATION #superfamily tobacco etch virus genome polyprotein; DEAD/H box helicase homology
KEYWORDS ATP; coat protein; genome-linked protein; inclusion protein; nucleus; P-loop; phosphoprotein; polyprotein
FEATURE
1-362 #product P1 protein #status predicted #label P1P\
363-820 #product helper component protein #status predicted #label HCP\
821-1175 #product P3 protein #status predicted #label P3P\
1176-1227 #product p6K1 protein #status predicted #label P6P\
1228-1870 #product cytoplasmic inclusion protein #status predicted #label CIP\
1313-1586 #domain DEAD/H box helicase homology #label DEAD\
1313-1320 #region nucleotide-binding motif A (P-loop)\
1398-1403 #region nucleotide-binding motif B\
1402-1405 #region DEXH motif\
1871-1923 #product p6K2 protein #status predicted #label PKP\
1924-2115 #product VPg protein #status predicted #label VPg\

2116-2358 #product nuclear inclusion a protein #status predicted #label NIA\

2359-2875 #product nuclear inclusion b protein #status predicted #label NIB\

2876-3163 #product coat protein #status predicted #label CAP\

1986 #binding_site phosphoryl-RNA (Tyr) (covalent) #status predicted

SUMMARY #length 3163 #molecular-weight 357819 #checksum 4441

Query Match 52.2%; Score 51; DB 1; Length 3163;

Best Local Similarity 41.7%; Pred. No. 4.61e+00;

Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 1299 GKIAHESDXDIL 1310

QY 1 GQIGNDPNRDIL 12

RESULT 12

ENTRY

TITLE

ORGANISM

DATE

ACCESSIONS

REFERENCE

#authors

A69672 #type complete

aspartate 1-decarboxylase pand - Bacillus subtilis

#formal_name Bacillus subtilis

05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 24-Sep-1998

A69672

A69580

Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Polotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kumano, M.; X.; Koetter, P.; Koningstein, G.; Krogh, S.; Lauber, J.; Kurita, K.; Lapidus, A.; Lardinois, S.; Li, H.; Masuda, S.; Lazarevic, V.; Lee, S.M.; Levine, A.; Mellado, R.P.; Mizuno, Maueel, C.; Medique, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256

#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

#cross-references MUID:98044033

#accession A69672

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-127 #label KUN

#cross-references GB:Z99115; GB:AL009126; NID:g2634478; PID:el183686;

#experimental_source strain 168

GENETICS

#gene

CLASSIFICATION

SUMMARY

Query Match 61.0%; Score 50; DB 2; Length 127;

Best Local Similarity 60.0%; Pred. No. 7.43e+00;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 118 LGNEPARTIL 127

QY 3 IGNDPNRDIL 12

RESULT 13

ENTRY

TITLE

ORGANISM

DATE

ACCESSIONS

REFERENCE

#authors

B70956 #type complete

GTP cyclohydrolase I (EC 3.5.4.16) - Mycobacterium tuberculosis (strain H37RV)

#formal_name Mycobacterium tuberculosis

17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Nov-1998

B70956

A70500

Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eigimeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544

#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references MUID:98295987

#accession B70956

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-202 #label COL

#cross-references GB:Z95557; GB:AL123456; NID:g3242276; PID:e316840;

PID:g2113967

#experimental_source strain H37RV

GENETICS

#gene

CLASSIFICATION

KEYWORDS

SUMMARY

Query Match 61.0%; Score 50; DB 2; Length 202;

Best Local Similarity 70.0%; Pred. No. 7.43e+00;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 32 IGEDPDRDGL 41

QY 3 IGNDPNRDIL 12

RESULT 14

ENTRY

TITLE

ALTERNATE_NAMES

ORGANISM

DATE

ACCESSIONS

REFERENCE

ISBYSS #type complete

protein disulfide-isomerase (EC 5.3.4.1) precursor - yeast (Saccharomyces cerevisiae)

protein YCL043c; protein YCL313; S-S rearrangase; thioredoxin-related glycoprotein 1; thyroid hormone-binding protein

#formal_name Saccharomyces cerevisiae

30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 12-Jun-1998

JX0182; PS0224; A39376; S15050; JS0634; S40913; A41713; S19372; S25349

JX0182

```
#authors      Tachikawa, H.; Miura, T.; Katakura, Y.; Mizunaga, T.
#journal      J. Biochem. (1991) 110:306-313
#title        Molecular structure of a yeast gene, PD11, encoding protein
               disulfide isomerase that is essential for cell growth.
#cross-references MUID:92105067
#accession    JX0182
               ##molecule_type DNA
               ##residues      1-522 ##label TAC
               ##cross-references GB:D00842; NID:g218506; PID:d1001182; PID:g218507
#accession    PS0224
               ##molecule_type protein
               ##residues      99-112;'X',185-193;211-217;'X',233-237;'X',239;284-286,
               'X',288-298;309-315;325-335 ##label TA2
REFERENCE
#authors      LaMantia, M.; Miura, T.; Tachikawa, H.; Kaplan, H.A.;
               Lennarz, W.J.; Mizunaga, T.
#journal      Proc. Natl. Acad. Sci. U.S.A. (1991) 88:4453-4457
#title        Glycosylation site binding protein and protein disulfide
               isomerase are identical and essential for cell viability in
               yeast.
#cross-references MUID:91239586
#accession    A39376
               ##molecule_type DNA
               ##residues      1-522 ##label LAM
               ##cross-references GB:M62815
REFERENCE
#authors      Scherens, B.; Dubois, E.; Messenguy, F.
#journal      Yeast (1991) 7:185-193
#title        Determination of the sequence of the yeast YCL313 gene
               localized on chromosome III. Homology with the protein
               disulfide isomerase (PDI gene product) of other organisms.
#cross-references MUID:91289690
#accession    S15050
               ##molecule_type DNA
               ##residues      1-522 ##label SCH
               ##cross-references EMBL:X57712; NID:g4801; PID:g4802
REFERENCE
#authors      Farquhar, R.; Honey, N.; Murant, S.J.; Bossier, P.; Schultz,
               L.; Montogomery, D.; Ellis, R.W.; Freedman, R.B.; Tuite,
               M.F.
#journal      Gene (1991) 108:81-89
#title        Protein disulfide isomerase is essential for viability in
               Saccharomyces cerevisiae.
#cross-references MUID:92104510
#accession    JS0634
               ##molecule_type DNA
               ##residues      1-113;'R',115-505;'EADAAEEA',506-522 ##label FAR
               ##cross-references EMBL:X54535; NID:g4119; PID:g4120
REFERENCE
#authors      Kuentzel, H.
#submission    submitted to the EMBL Data Library, April 1990
#accession    S40913
               ##molecule_type DNA
               ##residues      1-82;'V',84-142;'S',144,146-167;'E',169-225;'V',227-457,
               'S',459-505;'EADAAEEA',506-522 ##label KUE
               ##cross-references EMBL:X52313; NID:g3948; PID:g3949
REFERENCE
#authors      Guenther, R.; Braeuer, C.; Janetzky, B.; Foerster, H.H.;
               Ehbrecht, I.M.; Lehle, L.; Kuentzel, H.
#journal      J. Biol. Chem. (1991) 266:24557-24563
#title        The Saccharomyces cerevisiae TRG1 gene is essential for
               growth and encodes a luminal endoplasmic reticulum
               glycoprotein involved in the maturation of vacuolar
               carboxypeptidase.
#cross-references MUID:92105117
#accession    A41713
               ##molecule_type DNA
               ##residues      1-82;'V',84-142;'S',144,146-167;'E',169-225;'V',227-457,
               'S',459-505;'EADAAEEA',506-522 ##label GUE
               ##cross-references GB:M76982; NID:g173023; PID:g173024
REFERENCE
#authors      Dubois, E.; Pierard, A.; Gigot, D.; Glansdorff, N.;
               Messenguy, F.; Scherens, B.
```

```
#submission    submitted to the Protein Sequence Database, March 1992
#accession    S19372
               ##molecule_type DNA
               ##residues      1-522 ##label DUB
               ##cross-references EMBL:X59720; NID:g1907116; PID:e264537; PID:g5320;
               MIPS:YCL043C
REFERENCE
#authors      Scherens, B.; Messenguy, F.; Gigot, D.; Dubois, E.
#journal      Yeast (1992) 8:577-586
#title        The complete sequence of a 9,543 bp segment on the left arm
               of chromosome III reveals five open reading frames
               including glucokinase and the protein disulfide isomerase.
#accession    S25349
               ##molecule_type DNA
               ##residues      1-522 ##label SC2
               ##cross-references EMBL:X59720; EMBL:S43845; EMBL:S49180; EMBL:S58084;
               EMBL:S93798; NID:g1907116; PID:e264537; PID:g5320
GENETICS
#gene         SGD:PD11; MFp1; TRG1
               ##cross-references SGD:S0000548; MIPS:YCL043C
               ##map_position 3L
FUNCTION
#description   catalyzes rearrangement of both intrachain and interchain
               disulfide bonds
CLASSIFICATION
#superfamily   protein disulfide-isomerase; thioredoxin
               homology
KEYWORDS
#duplication;  endoplasmic reticulum; glycoprotein; homodimer;
               intramolecular oxidoreductase; isomerase; redox-active
               disulfide
FEATURE
1-22          #domain signal sequence #status predicted #label SIG\
23-522        #product protein disulfide-isomerase #status predicted
               #label MA1\
40-125        #domain thioredoxin homology #label TX1\
384-470        #domain thioredoxin homology #label TX2\
519-522        #region endoplasmic reticulum retention signal\
61-64,406-409  #disulfide_bonds redox-active #status predicted\
82,117,155,174,
425           #binding_site carbohydrate (Asn) (covalent) #status
               predicted\
90-97         #disulfide_bonds #status predicted
SUMMARY
#length 522 #molecular-weight 58227 #checksum 9937
Query Match      61.0%; Score 50; DB 1; Length 522;
Best Local Similarity 54.5%; Pred. No. 7.43e+00;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 388 EIVNDPKKDV1 398
               :| | | | :| |
QY 2 QIGNDPNRDIL 12
RESULT 15
ENTRY      S45177 #type complete
TITLE      BEM4 protein - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES
            protein P2561; protein YPL161c
ORGANISM   #formal_name Saccharomyces cerevisiae
            03-May-1994 #sequence_revision 02-Aug-1994 #text_change
            06-Feb-1998
ACCESSIONS S45177; S65172; S69438
REFERENCE   S45177
#authors    Mack, D.; Bender, A.
#submission submitted to the EMBL Data Library, January 1994
#accession  S45177
               ##molecule_type DNA
               ##residues      1-633 ##label MAC
               ##cross-references EMBL:L27816; NID:g450308; PID:g450309
REFERENCE   S65154
#authors    Purnelle, B.; Coster, F.; Goffeau, A.
#submission submitted to the Protein Sequence Database, May 1996
#accession  S65172
               ##molecule_type DNA
               ##residues      1-633 ##label PUR
```

```
##cross-references EMBL:Z73517; NID:gl370341; PID:e247048; PID:gl370342;
MIPS:YPL161c
##experimental_source strain S288C (AB972)
REFERENCE S69428
#authors Purnelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau,
A.
#submission submitted to the EMBL Data Library, March 1996
#description The sequence of 55 kb on the left arm of yeast chromosome XVI
identifies 28 open reading frames including 18 unknown
among which a new putative serine/threonine protein kinase,
a homologue to the human phosphotyrosyl phosphatase
activator PTPA and a homologue to the plant pleiotropic
regulator PRL1 of PPI and PP2a phosphatases.
#accession S69438
#molecule_type DNA
#residues 1-633 #label PUW
##cross-references EMBL:X96770; NID:gl403537; PID:e239041; PID:gl403548
GENETICS
#gene SGD:BEM4
##cross-references SGD:S0006082; MIPS:YPL161c
#map_position 16L
KEYWORDS transmembrane protein
FEATURE
387-403 #domain transmembrane #status predicted #label TMM
SUMMARY #length 633 #molecular-weight 70992 #checksum 3681
Query Match 61.0%; Score 50; DB 2; Length 633;
Best Local Similarity 70.0%; Pred. No. 7.43e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Db 102 IDNDDNREIL 111
| || |||
QY 3 IGNDPNRDIL 12
```

Search completed: Wed Aug 4 15:27:59 1999
Job time : 8 secs.

MPsrch_pp

protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:28:17 1999; MasPar time 3.14 Seconds

Tabular output not generated.

Title:

>US-09-049-847-3

Description:

{1-12} from US09049847.pep

Perfect Score:

82

Sequence:

1 GQIGNDPNRDIL 12

Scoring table:

PAM 150

Gap 15

Searched:

77977 seqs, 28268293 residues

Post-processing:

Minimum Match 0%

Listing first 45 summaries

Database:

swiss-prot37

1:swissprot

Statistics:

Mean 24.484; Variance 25.693; scale 0.953

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Release 3.1A

John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993-1998 University of Edinburgh, U.K.

Distribution rights by Oxford Molecular Ltd

(TM)

24	50	61.0	633	1	BEM4_YEAST	BEM4 PROTEIN.	2.71e+00
25	50	61.0	3770	1	ACVS_EMENI	DELTA-(L-ALPHA-AMINOAD	2.71e+00
26	49	59.8	841	1	CSW_DROME	PROTEIN-TYROSINE PHOSP	4.56e+00
27	48	58.5	179	1	GCHL_CAMJE	GTP CYCLOHYDROLASE I (7.58e+00
28	48	58.5	261	1	YEV6_YEAST	HYPOTHETICAL 29.7 KD P	7.58e+00
29	48	58.5	500	1	NU4C_TOBAC	NADH-PLASTOQUINONE OXI	7.58e+00
30	48	58.5	802	1	PUR2_YEAST	PHOSPHORIBOSYLAMINE--G	7.58e+00
31	48	58.5	1097	1	IMB3_HUMAN	IMPORTIN BETA-3 SUBUNI	7.58e+00
32	47	57.3	84	1	UNGI_BPPB2	URACIL-DNA GLYCOSYLASE	1.25e+01
33	47	57.3	378	1	KLAB_ECOLI	KLAA PROTEIN (TELA PRO	1.25e+01
34	47	57.3	514	1	TFEB_HUMAN	TFEB PROTEIN (FRAGMENT	1.25e+01
35	47	57.3	796	1	DHG_ECOLI	GLUCOSE DEHYDROGENASE	1.25e+01
36	47	57.3	821	1	SPK1_YEAST	PROTEIN KINASE SPK1 (E	1.25e+01
37	47	57.3	900	1	KDPD_CLOAB	SENSOR PROTEIN KDPD (E	1.25e+01
38	47	57.3	1093	1	PI4K_DICDI	PHOSPHATIDYLINOSITOL 4	1.25e+01
39	47	57.3	1379	1	MET_MOUSE	HEPATOCYTE GROWTH FACT	1.25e+01
40	47	57.3	1772	1	MSPL_PLAYO	MEROZOITE SURFACE PROT	1.25e+01
41	46	56.1	397	1	TRPB_HAEIN	TRYPTOPHAN SYNTHASE BE	2.04e+01
42	46	56.1	417	1	CBPB_HUMAN	CARBOXYPEPTIDASE B PRE	2.04e+01
43	46	56.1	655	1	CDST_KLEPN	CYCLOMALIODEXTRIN GLUC	2.04e+01
44	46	56.1	867	1	EF2_BLAHO	ELONGATION FACTOR 2 (E	2.04e+01
45	46	56.1	1184	1	BIMC_EMENI	KINESIN-LIKE PROTEIN B	2.04e+01

ALIGNMENTS

RESULT	1						
ID	TETX_CLOTE	STANDARD;	PRT;	1314	AA.		
AC	P04958;						
DI	13-AUG-1987 (REL. 05, CREATED)						
DI	13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)						
DI	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)						
DE	TETANUS TOXIN PRECURSOR (EC 3.4.24.68) (TENTOXYLYSIN).						
OS	CLOSTRIDIUM TETANI.						
OG	PLASMID.						
OC	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; CLOSTRIDIACEAE;						
OC	CLOSTRIDIUM.						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE; 87053814.						
RA	EISEL U., JARAUSCH W., GORETZKI K., HENSCHEN A., ENGELS J.,						
RA	WELLER U., HUDEL M., HABERMANN E., NIEMANN H.;						
RT	"Tetanus toxin: primary structure, expression in E. coli, and						
RT	homology with botulinum toxins.";						
RL	EMBO J. 5:2495-2502(1986).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=CN3911;						
RX	MEDLINE; 87040747.						
RA	FAIRWEATHER N.F., LYNNESS V.A.;						
RT	"The complete nucleotide sequence of tetanus toxin.";						
RL	NUCLEIC ACIDS RES. 14:7809-7812(1986).						
RN	[3]						
RP	SEQUENCE OF 742-1314 FROM N.A.						
RX	MEDLINE; 86085672.						
RA	FAIRWEATHER N.F., LYNNESS V.A., PICKARD D.J., ALLEN G., THOMSON R.O.;						
RT	"Cloning, nucleotide sequencing, and expression of tetanus toxin						
RT	fragment C in Escherichia coli.";						
RL	J. BACTERIOL. 165:21-27(1986).						
RN	[4]						
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.						
RX	MEDLINE; 90201034.						
RA	KRIEGLSTEIN K., HENSCHEN A., WELLER U., HABERMANN E.;						
RT	"Arrangement of disulfide bridges and positions of sulfhydryl groups						
RT	in tetanus toxin.";						
RL	EUR. J. BIOCHEM. 188:39-45(1990).						
RN	[5]						
RP	PARTIAL SEQUENCE.						
RX	MEDLINE; 92037649.						
RA	KRIEGLSTEIN K.G., HENSCHEN A.H., WELLER U., HABERMANN E.;						
RT	"Limited proteolysis of tetanus toxin. Relation to activity and						
RT	identification of cleavage sites.";						
RL	EUR. J. BIOCHEM. 202:41-51(1991).						

[6]
 RN IDENTIFICATION AS ZINC-PROTEASE.
 RP MEDLINE; 93010948.
 RA SCHIAVO G., POULAIN B., ROSSETTO O., BENFENATI F., TAUC L.,
 RA MONTECUCCO C.;
 RT "Tetanus toxin is a zinc protein and its inhibition of
 RL neurotransmitter release and protease activity depend on zinc.";
 RMBO J. 11:3577-3583(1992).
 RN [7]
 RP IDENTIFICATION OF SUBSTRATE.
 RX MEDLINE; 93063293.
 RA SCHIAVO G., BENFENATI F., POULAIN B., ROSSETTO O., DE LAURETO P.P.,
 RA DASGUPTA B.R., MONTECUCCO C.;
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release
 RT by proteolytic cleavage of synaptobrevin.";
 RL NATURE 359:832-835(1992).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.
 RX MEDLINE; 97475217.
 RA UMLAND I.C., WINGERT L.M., SWAMINATHAN S., FUREY W.F., SCHMIDT J.J.,
 RA SAX M.;
 RT "Structure of the receptor binding fragment HC of tetanus
 RT neurotoxin.";
 RL NAT. STRUCT. BIOL. 4:788-792(1997).
 CC -!- FUNCTION: TETANUS TOXIN ACIS BY INHIBITING NEUROTRANSMITTER
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77
 CC BOND OF SYNAPTOSREVIN-2.
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN
 CC SYNAPTOSREVIN.
 CC -!- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE
 CC AND ARE NON-TOXIC AFTER SEPARATION.
 CC -!- THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO GANGLIOSIDE RECEPTORS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE TETANUS/BOTULINUM NEUROTOXIN
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X04436; G40770; -.
 DR EMBL; M12739; G144921; -.
 DR EMBL; X06214; G40774; -.
 DR PIR; A25689; BTCLIN.
 DR PDB; 1AF9; 29-APR-98.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR NEUROTOXIN; TRANSMEMBRANE; HYDROLASE; METALLOPROTEASE; ZINC; PLASMID;
 KW 3D-STRUCTURE.
 FT INITMET 0 0
 FT CHAIN 1 456 TETANUS TOXIN LIGHT CHAIN.
 FT CHAIN 457 1314 TETANUS TOXIN HEAVY CHAIN.
 FT ACT_SITE 232 233 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 233 233 BY SIMILARITY.
 FT METAL 236 236 ZINC (CATALYTIC) (BY SIMILARITY).
 FT TRANSMEM 226 246 POTENTIAL.
 FT TRANSMEM 669 689 POTENTIAL.
 FT DISULFID 438 466 INTERCHAIN.
 FT DISULFID 1076 1092
 SQ SEQUENCE 1314 AA; 150550 MW; 26190E3E CRC32;

Query Match 100.0%; Score 82; DB 1; Length 1314;
 Best Local Similarity 100.0%; Pred. No. 7.41e-09;
 Matches. 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1272 GQIGNDPNRDIL 1283
 |||||
 QY 1 GQIGNDPNRDIL 12
 RESULT 2
 ID PRS6_DICDI STANDARD; PRT; 403 AA.
 AC P34123;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG (TAT-BINDING PROTEIN
 DE HOMOLOG 2).
 DE TBPB OR TBP2.
 OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
 OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE; 93312322.
 RA SHAW D.R., ENNIS H.L.;
 RT "Molecular cloning and developmental regulation of Dictyostelium
 RT discoideum homologues of the human and yeast HIV1 Tat-binding
 RT protein.";
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 193:1291-1296(1993).
 CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
 CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
 CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
 CC 26S COMPLEX (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
 CC -!- DEVELOPMENTAL STAGE: HIGHEST EXPRESSION IN VEGETATIVELY GROWING
 CC CELLS. THE LEVEL OF EXPRESSION FALLS STEADILY THROUGHOUT
 CC MULTICELLULAR DEVELOPMENT AND ARE NOT FOUND IN DORMANT OR
 CC GERMINATING SPORES.
 CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L16578; G290055; -.
 DR PIR; JN0611; JN0611.
 DR DICTYDB; DD01052; TBPB.
 DR PROSITE; PS00674; AAA; 1.
 DR PFAM; PF00004; AAA; 1.
 KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
 FT NP_BIND 191 198 ATP (POTENTIAL).
 SQ SEQUENCE 403 AA; 45542 MW; F1A7C9A7 CRC32;
 Query Match 63.4%; Score 52; DB 1; Length 403;
 Best Local Similarity 63.6%; Pred. No. 9.41e-01;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 178 QIGIDPPRGVL 188
 |||||
 QY 2 QIGNDPNRDIL 12
 RESULT 3
 ID PRS6_SOLTU STANDARD; PRT; 413 AA.
 AC P54778;
 DT 01-OCT-1996 (REL. 34, CREATED)
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.
 OS SOLANUM TUBEROSUM (POTATO).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; SOLANANAE; SOLANALES; SOLANACEAE; SOLANUM.


```
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. SUPERIOR;
RA HART J.K.; HANNAPEL D.J.;
RT "Isolation of an AAA superfamily cDNA clone from potato.";
RL (IN) PLANT GENE REGISTER PGR96-039.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U43398; G1155334; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 201 208 ATP (POTENTIAL).
SQ SEQUENCE 413 AA; 46532 MW; C75308E2 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 413;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 188 QIGIDPPRGVL 198
||| ||| :|
QY 2 QIGNDPNRDIL 12

RESULT 4
ID PRS6_CAEEL STANDARD; PRT; 414 AA.
AC P46502;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PROBABLE 26S PROTEASE REGULATORY SUBUNIT 6B.
GN F23F12.6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA DU Z.;
RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; J12965; G529215; -.
DR WORMPEP; F23F12.6; CE01253.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW HYPOTHETICAL PROTEIN; PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
```

```
FT NP_BIND 202 209 ATP (POTENTIAL).
SQ SEQUENCE 414 AA; 46358 MW; E1F26490 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 414;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 189 QIGIDPPRGVL 199
||| ||| :|
QY 2 QIGNDPNRDIL 12

RESULT 5
ID PRS6_RAT STANDARD; PRT; 418 AA.
AC Q63570;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (IAT-BINDING PROTEIN-7) (TBP-7).
GN PSMC4.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=LIVER;
RX MEDLINE; 96183075.
RA MAKINO Y., YOGOSAWA S., KANEMAKI M., YOSHIDA T., YAMANO K.,
RA KISHIMOTO T., MONCOLLIN V., EGLY J.M., MURAMATSU M., TAMURA T.;
RT "Structures of the rat proteasomal ATPases: determination of highly
RT conserved structural motifs and rules for their spacing.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 220:1049-1054(1996).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D50695; G1395186; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 206 213 ATP (POTENTIAL).
SQ SEQUENCE 418 AA; 47408 MW; 06A3A60B CRC32;

Query Match 63.4%; Score 52; DB 1; Length 418;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGIDPPRGVL 203
||| ||| :|
QY 2 QIGNDPNRDIL 12

RESULT 6
ID PRS6_MOUSE STANDARD; PRT; 418 AA.
AC P54775;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (CIP21).
GN PSMC4.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
```

```
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA CHOI H.S., SEOL W., MOORE D.D.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX. INTERACTS WITH THE ORPHAN NUCLEAR HORMONE RECEPTOR
CC MB67.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L76223; G1196528; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
DR PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 206 213 ATP (POTENTIAL).
SQ SEQUENCE 418 AA; 47281 MW; 8DCA5892 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 418;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGIDPPRGVL 203
QY 2 QIGNDPNRDIL 12

RESULT 7
ID PRS6_HUMAN STANDARD; PRT; 418 AA.
AC P43686;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B (TAT-BINDING PROTEIN-7) (TBP-7).
GN PSMC4 OR TBP7.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93126329.
RA OHANA B., MOORE P.A., RUBEN S.M., SOUTHGATE C.D., GREEN M.R.,
RA ROSEN C.A.;
RT "The type 1 human immunodeficiency virus Tat binding protein is a
RT transcriptional activator belonging to an additional family of
RT evolutionarily conserved genes.";
RL PROC. NATL. ACAD. SCI. U.S.A. 90:138-142(1993).
RN [2]
RP PARTIAL SEQUENCE, REVISION TO C-TERMINAL, AND FUNCTION.
RX MEDLINE; 94338582.
RA DUBIEL W., FERRELL K., RECHSTEINER M.;
RT "Tat-binding protein 7 is a subunit of the 26S protease.";
RL BIOL. CHEM. HOPPE-SEYLER 375:237-240(1994).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR.
CC -!- PTM: THE N-TERMINAL IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
DR MM; 602707; -.
DR PROSITE; PS00674; AAA; 1.
```

```
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 206 213 ATP (POTENTIAL).
SQ SEQUENCE 418 AA; 47336 MW; 4D06C2AF CRC32;

Query Match 63.4%; Score 52; DB 1; Length 418;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 193 QIGIDPPRGVL 203
QY 2 QIGNDPNRDIL 12

RESULT 8
ID PRS6_ASPNG STANDARD; PRT; 423 AA.
AC P78578;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.
GN TBPA.
OS ASPERGILLUS NIGER.
OC EUKARYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; PLECTOMYCETES;
OC EUROTIALES; TRICHOCOMACEAE; MITOSPORIC TRICHOCOMACEAE; ASPERGILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 120.49 / N400;
RA DELEU C., JARAI G., BUXTON F.;
RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U15601; G1777414; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 207 214 ATP (POTENTIAL).
SQ SEQUENCE 423 AA; 47222 MW; 14E49380 CRC32;

Query Match 63.4%; Score 52; DB 1; Length 423;
Best Local Similarity 63.6%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 194 QIGIDPPRGVL 204
QY 2 QIGNDPNRDIL 12

RESULT 9
ID PRS6_YEAST STANDARD; PRT; 428 AA.
AC P33298;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG (YNT1 PROTEIN) (IAT-BINDING
DE HOMOLOG 2).
GN YTA2 OR YNT1 OR YDR394W OR D9509.14.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C836;
RX MEDLINE; 95274317.
RA SCHNALL R., MANNHAUPT G., STUCKA R., TAUER R., EHNLE S.,
RA SCHWARZLOSE C., VETTER I., FELDMANN H.;
RT "Identification of a set of yeast genes coding for a novel family of
RT putative ATPases with high similarity to constituents of the 26S
RT protease complex.";
RL YEAST 10:1141-1155(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA CAMPBELL C.L., TANAKA N., WHITE K.H., THORSNESS P.E.;
RL SUBMITTED (FEB-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RA DIETRICH F.S., MULLIGAN J., ALLEN E., ARAUJO R., AVILES E.,
RA BERNI A., CARPENTER J., CHEN E., CHERRY J.M., CHUNG E., DUNCAN M.,
RA HUNICKE-SMITH S., HYMAN R., KOMP C., LASHKARI D., LEW H., LIN D.,
RA MOSEDALE D., NAKAHARA K., NAMATH A., OEFNER P., OH C., PETEL F.X.,
RA ROBERTS D., SCHRAMM S., SCHROEDER M., SHOGREN T., SHROFF N.,
RA WINANT A., YELTON M., BOTSTEIN D., DAVIS R.W.;
RL SUBMITTED (JUL-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF URQUITTINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X73570; G313880; -.
CC DR EMBL; U06229; G458389; -.
CC DR EMBL; U32274; G927327; -.
CC DR PIR; S46606; S46606.
CC DR SGD; L0002556; YTA2.
CC DR PROSITE; PS00674; AAA; 1.
CC DR PFAM; PF00004; AAA; 1.
CC DR PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
KW NP_BIND 213 220 ATP (POTENTIAL).
FT CONFLICT 342 342 Y -> S (IN REF. 3).
SQ SEQUENCE 428 AA; 47969 MW; E96A890E CRC32;
Query Match 63.4%; Score 52; DB 1; Length 428;
Best Local Similarity 63.5%; Pred. No. 9.41e-01;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Db 200 QIGIDPPRGVL 210
QY 2 QIGNDPNRDIL 12
RESULT 10
ID POLG.YEFV1 STANDARD; PRI; 3411 AA.
AC P03314; O42028;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
DE RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS YELLOW FEVER VIRUS (STRAIN 17D).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
```

```

RP SEQUENCE FROM N.A.
RX MEDLINE; 85272570.
RA RICE C.M., LENCHES E.M., EDDY S.R., SHIN S.J., SHEETS R.L.,
RA STRAUSS J.H.;
RT "Nucleotide sequence of yellow fever virus: implications for
RT flavivirus gene expression and evolution.";
RL SCIENCE 229:726-733(1985).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; X03700; G59339; -.
CC DR EMBL; K02749; G336193; -.
CC DR PIR; A03914; GNWVY.
CC DR PFAM; PF00869; Flavi_glycoprot; 1.
CC DR PFAM; PF00948; Flavi_NS1; 1.
CC DR PFAM; PF00949; Flavi_helicase; 1.
CC DR PFAM; PF00972; Flavi_NS5; 1.
CC DR PFAM; PF01002; Flavi_NS2B; 1.
CC DR PFAM; PF01003; Flavi_capsid; 1.
CC DR PFAM; PF01004; Flavi_M; 1.
CC DR PFAM; PF01005; Flavi_NS2A; 1.
CC DR HSP; P14336; LSVB.
KW POLYPROTEIN; GLYCOPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE;
KW CORE PROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING;
KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CHAIN 1 121 CELLULAR AMINOPEPTIDASE.
FT PROPEP 122 210 CAPSID PROTEIN C.
FT CHAIN 211 285 ENVELOPE GLYCOPROTEIN M.
FT CHAIN 286 778 MAJOR ENVELOPE PROTEIN E.
FT CHAIN 779 1130 NONSTRUCTURAL PROTEIN NS1.
FT CHAIN 1131 1354 NONSTRUCTURAL PROTEIN NS2A.
FT CHAIN 1355 1484 NONSTRUCTURAL PROTEIN NS2B.
FT CHAIN 1485 2107 HELICASE (NS3).
FT CHAIN 2108 2256 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 2257 2506 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2507 3411 RNA-DIRECTED RNA POLYMERASE (NS5).
FT TRANSMEM 249 269 POTENTIAL.
FT TRANSMEM 271 285 POTENTIAL.
FT TRANSMEM 740 753 POTENTIAL.
FT TRANSMEM 755 778 POTENTIAL.
FT TRANSMEM 1159 1180 POTENTIAL.
FT DOMAIN 383 396 INVOLVED IN FUSION.
FT NP_BIND 1682 1776 ATP (POTENTIAL).
FT SITE 1773 1776 DEAH BOX.
FT DISULFID 288 315 BY SIMILARITY.
FT DISULFID 345 401 BY SIMILARITY.
FT DISULFID 359 390 BY SIMILARITY.
FT DISULFID 377 406 BY SIMILARITY.
FT DISULFID 467 568 BY SIMILARITY.
FT DISULFID 585 615 BY SIMILARITY.
FT CARBOHYD 134 134 POTENTIAL.
FT CARBOHYD 150 150 POTENTIAL.
FT CARBOHYD 908 908 POTENTIAL.
FT CARBOHYD 986 986 POTENTIAL.
FT CARBOHYD 2320 2320 POTENTIAL.
FT CARBOHYD 2346 2346 POTENTIAL.
FT CARBOHYD 2467 2467 POTENTIAL.
SQ SEQUENCE 3411 AA; 379512 MW; E85D316D CRC32;
```

```
Query Match          63.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 9.41e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGNPNRD 1956
    1 GQIGNDPNRD 10

RESULT 11
ID POLG_YEFV2          STANDARD;          PRT; 3411 AA.
AC P19901;
DT 01-FEB-1991 (REL. 17, CREATED)
DI 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DI 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
DE RNA POLYMERASE (EC 2.7.7.48) (NS5)].
OS YELLOW FEVER VIRUS (STRAIN PASTEUR 17D-204).
OC VIRUSES; SSRNA POSITIVE-SIRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 89282413.
RA DUPUY A., DESPRES P., CAHOUR A., GIRARD M., BOULOY M.;
RT "Nucleotide sequence comparison of the genome of two 17D-204 yellow
RT fever vaccines.";
RL NUCLEIC ACIDS RES. 17:3989-3989(1989).
CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRNA.
CC -----
CC This SWISS-PROI entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15062; -; NOT_ANNOTATED_CDS.
DR PIR; S07757; GNWVYP.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
DR HSSP; P14336; 1SVB.
KW POLYPROTEIN; GLYCOPROTEIN; TRANSFERASE; RNA-DIRECTED RNA POLYMERASE;
KW CORE PROTEIN; COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING;
KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.
FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CAPSID PROTEIN C.
FT CHAIN 1 121 ENVELOPE GLYCOPROTEIN M.
FT PROPEP 122 210 MAJOR ENVELOPE PROTEIN E.
FT CHAIN 211 285 NONSTRUCTURAL PROTEIN NS1.
FT CHAIN 286 778 NONSTRUCTURAL PROTEIN NS2A.
FT CHAIN 779 1130 NONSTRUCTURAL PROTEIN NS2B.
FT CHAIN 1131 1354 HELICASE (NS3).
FT CHAIN 1355 1484 NONSTRUCTURAL PROTEIN NS4A.
FT CHAIN 1485 2107 NONSTRUCTURAL PROTEIN NS4B.
FT CHAIN 2108 2256 RNA-DIRECTED RNA POLYMERASE (NS5).
FT CHAIN 2257 2506
FT CHAIN 2507 3411
```

```
NP_BIND 1682 1689 ATP (POTENTIAL).
SITE 1773 1776 DEAH BOX.
TRANSMEM 249 269 POTENTIAL.
TRANSMEM 271 285 POTENTIAL.
TRANSMEM 740 753 POTENTIAL.
TRANSMEM 755 778 POTENTIAL.
TRANSMEM 1159 1180 POTENTIAL.
DISULFID 288 315 BY SIMILARITY.
DISULFID 345 401 BY SIMILARITY.
DISULFID 359 390 BY SIMILARITY.
DISULFID 377 406 BY SIMILARITY.
DISULFID 467 568 BY SIMILARITY.
DISULFID 585 615 BY SIMILARITY.
CARBOHYD 134 134 POTENTIAL.
CARBOHYD 150 150 POTENTIAL.
CARBOHYD 908 908 POTENTIAL.
CARBOHYD 986 986 POTENTIAL.
CARBOHYD 2320 2320 POTENTIAL.
CARBOHYD 2346 2346 POTENTIAL.
CARBOHYD 2467 2467 POTENTIAL.
SEQUENCE 3411 AA; 379524 MW; ADCCF22B CRC32;

Query Match          63.4%; Score 52; DB 1; Length 3411;
Best Local Similarity 70.0%; Pred. No. 9.41e-01;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGNPNRD 1956
    1 GQIGNDPNRD 10

RESULT 12
ID GCH1_BACSU          STANDARD;          PRT; 190 AA.
AC P19465;
DT 01-FEB-1991 (REL. 17, CREATED)
DI 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DI 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE GTP CYCLOHYDROLASE I (EC 3.5.4.16) (GTP-CH-I).
MTRA.
OS BACILLUS SUBTILIS.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
OC BACILLUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91062353.
RA GOLLNICK P., ISHINO S., KURODA M.I., HENNER D.J., YANOFSKY C.;
RT "The mtr locus is a two-gene operon required for transcription
RT attenuation in the trp operon of Bacillus subtilis.";
RL PROC. NATL. ACAD. SCI. U.S.A. 87:8726-8730(1990).
RN [2]
RP SEQUENCE OF 1-19 FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE; 91216992.
RA MICKA B., GROCH N., HEINEMANN U., MARAHIEL M.A.;
RT "Molecular cloning, nucleotide sequence, and characterization of the
RT Bacillus subtilis gene encoding the DNA-binding protein HBSu.";
RL J. BACTERIOL. 173:3191-3198(1991).
RN [3]
RP FUNCTION.
RX MEDLINE; 92202128.
RA BABITZKE P., GOLLNICK P., YANOFSKY C.;
RT "The mtrAB operon of Bacillus subtilis encodes GTP cyclohydrolase I
RT (MtrA), an enzyme involved in folic acid biosynthesis, and MtrB, a
RT regulator of tryptophan biosynthesis.";
RL J. BACTERIOL. 174:2059-2064(1992).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE; 95194311.
RA DE SAIZIEU A., VANKAN P., VAN LOON A.P.;
RT "Enzymic characterization of Bacillus subtilis Gtp cyclohydrolase I.
RT Evidence for a chemical dephosphorylation of dihydroneopterin
RT triphosphate.";
RL BIOCHEM. J. 306:371-377(1995).
```


CC -!- CATALYTIC ACTIVITY: GTP + 2 H(2)O = FORMATE + 2-AMINO-4-HYDROXY-
CC 6-(ERYTHRO-1,2,3-TRIHYDROXYPROPYL)DIHYDROPTERIDINE TRIPHOSPHATE.
CC -!- ENZYME REGULATION: K+ IONS MODERATELY INCREASES THE VMAX., WHEREAS
CC UTP AND CA2+ AND MG2+ IONS DRASTICALLY INCREASE THE KM FOR GTP.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF TETRAHYDROFOLATE.
CC -!- SUBUNIT: HOMOPOLYMER.
CC -!- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M37320; G14323L; -.
DR EMBL; M80245; G143799; -.
DR EMBL; X52418; -; NOT_ANNOTATED_CDS.
DR PIR; A38256; A38256.
DR PIR; A39409; A39409.
DR SUBTILIST; BG10277; MTRA.
DR PROSITE; PS00859; GTP_CYCLOHYDROL_1_1; 1.
DR PROSITE; PS00860; GTP_CYCLOHYDROL_1_2; 1.
DR PFAM; PF01227; GTP_cyclohydroi; 1.
KW ONE-CARBON METABOLISM; HYDROLASE; ALLOSTERIC ENZYME.
FT DISULFID 78 150 BY SIMILARITY
SQ SEQUENCE 190 AA; 21219 MW; 2790F255 CRC32;

Query Match 62.2%; Score 51; DB 1; Length 190;
Best Local Similarity 70.0%; Pred. No. 1.60e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 20 IGEDPNREGL 29
||:||||:
QY 3 IGNDPNRDIL 12

RESULT 13
ID VMAT_NDVB STANDARD; PRT; 364 AA.
AC P06157;
DI 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
GN M.
OS NEWCASTLE DISEASE VIRUS (STRAIN BEAUDETTE C/45) (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87066775.
RA CHAMBERS P., MILLAR N.S., PLATT S.G., EMMERSON P.T.;
RT "Nucleotide sequence of the gene encoding the matrix protein of
RT Newcastle disease virus."
RL NUCLEIC ACIDS RES. 14:9051-9061(1986).
CC -!- FUNCTION: THE M PROTEIN HAS A CRUCIAL ROLE IN VIRUS ASSEMBLY
CC AND INTERACTS WITH THE RNP COMPLEX AS WELL AS WITH THE VIRAL
CC MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X04587; G60941; -.
DR PIR; A26111; MFNZNC.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN; ENVELOPE PROTEIN.
SQ SEQUENCE 364 AA; 39605 MW; 635391DD CRC32;

Query Match 62.2%; Score 51; DB 1; Length 364;
Best Local Similarity 41.7%; Pred. No. 1.60e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | :|:|:|
QY 1 QGIGNDPNRDIL 12

RESULT 14
ID GANA_PSEFL STANDARD; PRT; 376 AA.
AC P48841;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE ARABINOGALACTAN ENDO-1,4-BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.89)
DE (ENDO-1,4-BETA-GALACTANASE) (GALACTANASE).
GN GANA.

OS PSEUDOMONAS FLUORESCENS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC PSEUDOMONAS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP. CELLULOSA;
RA BRAITHWAITE K.L., GILBERT H.J.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-GALACTOSIDIC
CC LINKAGES IN ARABINOGALACTANS.
CC -!- SIMILARITY: BELONGS TO FAMILY 53 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91885; G1017437; -.
KW HYDROLASE; GLYCOSIDASE; SIGNAL.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 376 ARABINOGALACTAN ENDO-1,4-BETA-
FT GALACTOSIDASE.
SQ SEQUENCE 376 AA; 42315 MW; 56D4C50F CRC32;

Query Match 62.2%; Score 51; DB 1; Length 376;
Best Local Similarity 54.5%; Pred. No. 1.60e+00;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Db 157 QVGNETNIEIL 167
|:|:|:|:
QY 2 QIGNDPNRDIL 12

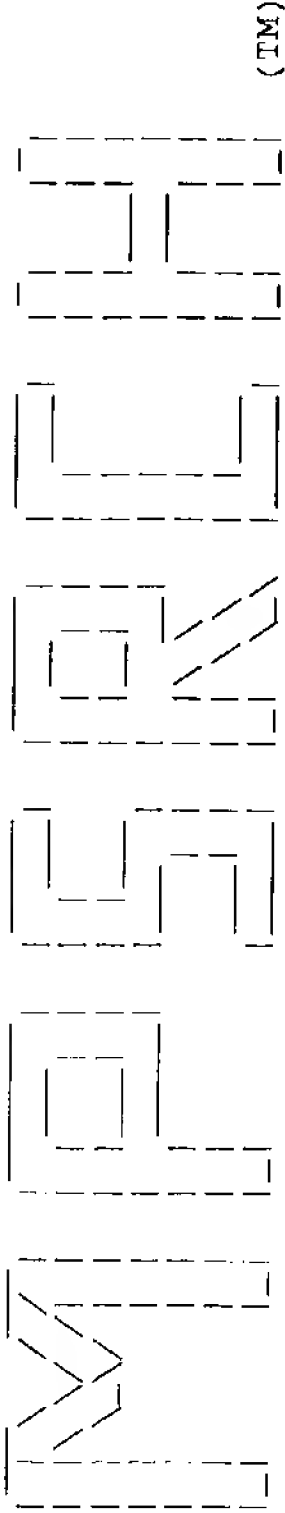
RESULT 15
ID PRS6_MANSE STANDARD; PRT; 415 AA.
AC P46507;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6S (ATPASE MS73).
OS MANDUCA SEXTA (TOBACCO HAWNMOTH) (TOBACCO HORNWORM).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; SPHINGIODEA; SPHINGIDAE; SPHINGINAE; MANDUCA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTERSEGMENTAL MUSCLE;
RX MEDLINE; 95130567.
RA DAWSON S.P., ARNOLD J., MAYER N.J., REYNOLDS S., BILLETT M.A.,
RA GORDON C., COLLEAUX L., KLOETZEL P., TANAKA K., MAYER R.J.;
RT "Developmental changes of the 26 S proteasome in abdominal
RT intersegmental muscles of Manduca sexta during programmed cell

RT death.";
RL J. BIOL. CHEM. 270:1850-1858(1995).
CC -!- FUNCTION: THE 26S PROTEASE IS INVOLVED IN THE ATP-DEPENDENT
CC DEGRADATION OF UBIQUITINATED PROTEINS. THE REGULATORY (OR ATPASE)
CC COMPLEX CONFERS ATP DEPENDENCY AND SUBSTRATE SPECIFICITY TO THE
CC 26S COMPLEX.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC AND NUCLEAR (POTENTIAL).
CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THE GENE DRAMATICALLY INCREASES
CC IN THE PRE-ECLOSION PERIOD.
CC -!- SIMILARITY: BELONGS TO THE AAA FAMILY OF ATPASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38135; G559486; -.
DR PROSITE; PS00674; AAA; 1.
DR PFAM; PF00004; AAA; 1.
KW PROTEASOME; ATP-BINDING; NUCLEAR PROTEIN.
FT NP_BIND 203 210 ATP (POTENTIAL).
SQ SEQUENCE 415 AA; 47053 MW; 7BD9FB35 CRC32;

Query Match 62.2%; Score 51; DB 1; Length 415;
Best Local Similarity 54.5%; Pred.No. 1.60e+00;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 190 QIGIEPPRGVL 200
|||:|:|
QY 2 QIGNDPNRDIL 12

Search completed: Wed Aug 4 15:28:22 1999
Job time : 5 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:28:40 1999; MasPar time 7.10 seconds
Tabular output not generated. 92.235 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrcmbl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 23.731; Variance 25.076; scale 0.946

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	56	68.3	473	2	051898	HYPOTHETICAL 51.9 KD P
2	52	63.4	389	3	074894	26S PROTEASE REGULATOR
3	52	63.4	3410	14	Q89278	POLYPROTEIN.
4	52	63.4	3410	14	Q89277	POLYPROTEIN.
5	52	63.4	3410	14	Q89275	POLYPROTEIN PRECURSOR.
6	52	63.4	3410	14	Q89276	POLYPROTEIN PRECURSOR.
7	52	63.4	3411	14	Q91857	POLYPROTEIN.
8	52	63.4	3411	14	Q98803	POLYPROTEIN.
9	51	62.2	158	2	Q59508	PREPILIN.
10	51	62.2	206	11	Q88874	CYCLIN K (FRAGMENT).
11	51	62.2	357	4	Q75909	CYCLIN K.
12	51	62.2	364	14	Q92330	MATRIX PROTEIN.
13	51	62.2	364	14	Q83834	MATRIX PROTEIN.
14	51	62.2	364	14	Q83838	MATRIX PROTEIN.
15	51	62.2	364	14	Q83840	MATRIX PROTEIN.
16	51	62.2	687	4	Q60331	KIAA0589 PROTEIN (FRAG
17	50	61.0	202	2	Q06273	GCHA.
18	50	61.0	253	5	Q17381	F52G3.2 PROTEIN.
19	49	59.8	184	2	Q66603	GIP CYCLOHYDROLASE I.
20	49	59.8	205	2	Q69531	GIP CYCLOHYDROLASE I.

21	49	59.8	400	2	068995	HYPOTHETICAL 44.2 KD P
22	49	59.8	584	10	004886	PECTINESTERASE (EC 3.1
23	49	59.8	584	10	004888	PECTINESTERASE (EC 3.1
24	49	59.8	613	5	Q23841	PUTATIVE GDH.
25	49	59.8	764	5	Q24708	CORKSCREW PROTEIN (EC
26	49	59.8	841	5	Q24032	CORKSCREW PROTEIN Y122
27	49	59.8	945	5	Q24033	CORKSCREW PROTEIN 4A (
28	49	59.8	1218	4	Q15816	TRANSMEMBRANE PROTEIN
29	49	59.8	1596	11	Q62766	SSECKS.
30	49	59.8	2374	5	Q45377	F22G12.5 PROTEIN.
31	48	58.5	144	1	Q28064	METHYLMALONYL-COA MUTA
32	48	58.5	194	2	Q33723	GTP CYCLOHYDROLASE (EC
33	48	58.5	230	14	Q82681	POLYMERASE-ASSOCIATED
34	48	58.5	296	2	P74223	GLUTAMINE-BINDING PROT
35	48	58.5	780	10	Q04346	REGULATORY PROTEIN VIV
36	47	57.3	209	14	Q40596	NEF GENE (STRAIN KU8)
37	47	57.3	210	3	P87174	PROBABLE DNA-BINDING P
38	47	57.3	299	2	Q53342	HYPOTHETICAL 31.6 KD P
39	47	57.3	341	5	Q20249	F40G12.1 PROTEIN.
40	47	57.3	364	14	Q90338	MATRIX PROTEIN.
41	47	57.3	562	2	P71654	HYPOTHETICAL 58.8 KD P
42	47	57.3	661	11	Q70161	PHOSPHATIDYLINOSITOL-4
43	47	57.3	751	2	Q33613	CATALASE (EC 1.11.1.6)
44	47	57.3	1105	4	Q60518	RAN-GTP BINDING PROTEI
45	47	57.3	2531	5	Q16004	NOTCH HOMOLOG.

ALIGNMENTS

RESULT 1
ID 051898 PRELIMINARY; PRT; 473 AA.
AC 051898;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 51.9 KD PROTEIN.
OS BORRELIA AFZELII.
OC BACTERIA; SPIROCHAETALES; SPIROCHAETACEAE; BORRELIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R-IP3;
RX MEDLINE; 98065594.
RA CASJENS S., MURPHY M., DELANGE M., SAMPSON L., VAN VUGT R.,
RA HUANG W.M.;
RT "Telomeres of the linear chromosomes of Lyme disease spirochaetes:
RT nucleotide sequence and possible exchange with linear plasmid
RT telomeres.";
RL MOL. MICROBIOL. 26:581-596(1997).
DR EMBL; AF008219; G2697115; -.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 473 AA; 51868 MW; BFA2E4DE CRC32;

Query Match 68.3%; Score 56; DB 2; Length 473;
Best Local Similarity 54.5%; Pred. No. 1.36e-01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 35 QIGDGPKEII 46
|||:|:|:|:
QY 2 QIGNDPNRDIL 12

RESULT 2
ID 074894 PRELIMINARY; PRT; 389 AA.
AC 074894;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE 26S PROTEASE REGULATORY SUBUNIT 6B HOMOLOG.
GN SPC0576.10C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.

RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17DD;
RA GALLER R.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U17066; G829367; -.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN.
FT CHAIN 210 284 M PROTEIN.
FT CHAIN 285 777 ENVELOPE PROTEIN.
FT CHAIN 778 1186 NS1.
FT CHAIN 1187 1353 POTENTIAL.
FT CHAIN 1354 1482 POTENTIAL.
FT CHAIN 1483 2105 NS3.
FT CHAIN 2107 2393 POTENTIAL.
FT CHAIN 2394 2505 POTENTIAL.
FT CHAIN 2506 2506 NS5.
SQ SEQUENCE 3410 AA; 379301 MW; 06C3D225 CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3410;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1946 GRIGNPNRD 1955
|:|:|:|
QY 1 GQIGNDPNRD 10

RESULT 6
ID Q89276 PRELIMINARY; PRT; 3410 AA.
AC Q89276;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN PRECURSOR.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17D-213;
RA DOS SANTOS C.N., POST P.R., CARVALHO R., FERREIRA I., RICE C.M.,
RA GALLER R.;
RL VIRUS RES. 35:35-41(1994).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=VACCINE STRAIN 17D-213;
RA GALLER R.;
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U17067; G829369; -.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN.
FT CHAIN 210 284 M PROTEIN.
FT CHAIN 285 777 ENVELOPE PROTEIN.
FT CHAIN 778 1186 NS1.
FT CHAIN 1187 1353 PUTATIVE NON-STRUCTURAL PROTEIN NS2A.
FT CHAIN 1354 1482 PUTATIVE NON-STRUCTURAL PROTEIN NS2B.
FT CHAIN 1483 2105 NS3.
FT CHAIN 2107 2393 PUTATIVE NON-STRUCTURAL PROTEIN NS4A.
FT CHAIN 2394 2505 PUTATIVE NON-STRUCTURAL PROTEIN NS4B.

FT CHAIN 2506 2506 NS5.
SQ SEQUENCE 3410 AA; 379357 MW; D225424D CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3410;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1946 GRIGNPNRD 1955
|:|:|:|
QY 1 GQIGNDPNRD 10

RESULT 7
ID Q91857 PRELIMINARY; PRT; 3411 AA.
AC Q91857;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17D-204-USA VACCINE;
RX MEDLINE; 98376360.
RA XIE H., CASS A.R., BARRETT A.D.T.;
RT "Yellow fever 17D vaccine virus isolated from healthy vaccinees
RT accumulates very few mutations.";
RL VIRUS RES. 55:93-99(1998).
DR EMBL; AF052438; G3288892; -.
KW POLYPROTEIN.
SQ SEQUENCE 3411 AA; 379584 MW; 1C9F6704 CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3411;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1947 GRIGNPNRD 1956
|:|:|:|
QY 1 GQIGNDPNRD 10

RESULT 8
ID Q98803 PRELIMINARY; PRT; 3411 AA.
AC Q98803;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLYPROTEIN.
OS YELLOW FEVER VIRUS.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; FLAVIVIRIDAE;
OC FLAVIVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85-82H IVORY COAST;
RA PISANO M.R., TOLOU H., NICOLI J.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U54798; G1314775; -.
DR PFAM; PF00869; Flavi_glycoprot; 1.
DR PFAM; PF00948; Flavi_NS1; 1.
DR PFAM; PF00949; Flavi_helicase; 1.
DR PFAM; PF00972; Flavi_NS5; 1.
DR PFAM; PF01002; Flavi_NS2B; 1.
DR PFAM; PF01003; Flavi_capsid; 1.
DR PFAM; PF01004; Flavi_M; 1.
DR PFAM; PF01005; Flavi_NS2A; 1.
KW POLYPROTEIN.
SQ SEQUENCE 3411 AA; 379170 MW; E3416193 CRC32;

Query Match 63.4%; Score 52; DB 14; Length 3411;
Best Local Similarity 70.0%; Pred. No. 1.28e+00;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```
Db 1947 GRIGNPNRD 1956
QY 1 GQIGNDPNRD 10

RESULT 9
ID Q59508 PRELIMINARY; PRT; 158 AA.
AC Q59508;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PREPILIN.
OS MORAXELLA BOVIS.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; PSEUDOMONAS GROUP;
OC MORAXELLACEAE; MORAXELLA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3W07 / SEROGROUP B;
RX MEDLINE; 94327452.
RA ATWELL J.L., TENNENT J.M., LEPPER A.W., ELLEMAN T.C.;
RT "Characterization of pilin genes from seven serologically defined
RT prototype strains of Moraxella bovis.";
RL J. BACTERIOL. 176:4875-4882(1994).
DR EMBL; L32969; G488318; -.
DR PROSITE; PS00409; PROKAR_NTER_METHYL; 1.
DR PFAM; PF00114; pilin; 1.
KW METHYLATION.
FT MOD_RES 7 7 METHYLATION (BY SIMILARITY).
SQ SEQUENCE 158 AA; 16590 MW; D2841B31 CRC32;

Query Match 62.2%; Score 51; DB 2; Length 158;
Best Local Similarity 66.7%; Pred. No. 2.20e+00;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 109 GTIGNRANKDIL 120
QY 1 GQIGNDPNRDIL 12

RESULT 10
ID O88874 PRELIMINARY; PRT; 206 AA.
AC O88874;
DI 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYCLIN K (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98298273.
RA EDWARDS M.C., WONG C., ELLEDGE S.J.;
RT "Human cyclin K, a novel RNA polymerase II-associated cyclin
RT possessing both carboxy-terminal domain kinase and Cdk-activating
RT kinase activity.";
RL MOL. CELL. BIOL. 18:4291-4300(1998).
DR EMBL; AF060517; G3746552; -.
FT NON_TER 206 206
SQ SEQUENCE 206 AA; 24071 MW; A154D984 CRC32;

Query Match 62.2%; Score 51; DB 11; Length 206;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 114 GQFGDDPKKEVM 125
QY 1 GQIGNDPNRDIL 12

RESULT 11
ID O75909 PRELIMINARY; PRT; 357 AA.
```

```
AC O75909;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CYCLIN K.
GN CPR4.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98298273.
RA EDWARDS M.C., WONG C., ELLEDGE S.J.;
RT "Human cyclin K, a novel RNA polymerase II-associated cyclin
RT possessing both carboxy-terminal domain kinase and Cdk-activating
RT kinase activity.";
RL MOL. CELL. BIOL. 18:4291-4300(1998).
DR EMBL; AF060515; G3746549; -.
SQ SEQUENCE 357 AA; 41293 MW; 547C5CE6 CRC32;

Query Match 62.2%; Score 51; DB 4; Length 357;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Db 130 GQFGDDPKKEVM 141
QY 1 GQIGNDPNRDIL 12

RESULT 12
ID O92330 PRELIMINARY; PRT; 364 AA.
AC O92330;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
GN M.
OS NEWCASTLE DISEASE VIRUS STRAIN F48E9.
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F48E9;
RA CHEN H., LIU S., CHEN J., ZHANG B., LU J.;
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL; AF089819; G3695291; -.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39783 MW; B7CE8893 CRC32;

Query Match 62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
QY 1 GQIGNDPNRDIL 12

RESULT 13
ID Q83834 PRELIMINARY; PRT; 364 AA.
AC Q83834;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMBER;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).
```


Job time : 10 secs.

```
DR EMBL; U25831; G1016659; -.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39723 MW; F0E2A4AB CRC32;

Query Match      62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | :|:|:|
QY 1 GQIGNDPNRDIL 12

RESULT 14
ID Q83838 PRELIMINARY; PRI; 364 AA.
AC Q83838;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (IREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TEXAS/GB;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).
DR EMBL; U25835; G1016667; -.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39464 MW; 72A78A51 CRC32;

Query Match      62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | :|:|:|
QY 1 GQIGNDPNRDIL 12

RESULT 15
ID Q83840 PRELIMINARY; PRI; 364 AA.
AC Q83840;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE MATRIX PROTEIN.
OS NEWCASTLE DISEASE VIRUS (NDV).
OC VIRUSES; SSRNA NEGATIVE-STRAND VIRUSES; MONONEGAVIRALES;
OC PARAMYXOVIRIDAE; PARAMYXOVIRINAE; RUBULAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ULSTER;
RA SEAL B.S.;
RL VIRUS GENES 11:217-224(1996).
DR EMBL; U25837; G1016671; -.
DR PFAM; PF00661; Matrix; 1.
KW MATRIX PROTEIN.
SQ SEQUENCE 364 AA; 39645 MW; 4F9287D2 CRC32;

Query Match      62.2%; Score 51; DB 14; Length 364;
Best Local Similarity 41.7%; Pred. No. 2.20e+00;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 76 GMINDNPKRELL 87
| | :|:|:|
QY 1 GQIGNDPNRDIL 12

Search completed: Wed Aug 4 15:28:50 1999
```

This Page Blank (uspto)

[M][E][R][E] (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:27:21 1999; MasPar time 9.30 Seconds
Tabular output not generated. 27.440 Million cell updates/sec
Title: >US-09-049-847-3
Description: (1-12) from US09049847.pap
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12
Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39
Statistics: Mean 17.007; Variance 46.283; scale 0.367
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				%			
Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	82	100.0	12	36	W67035	Tetanus toxin fragmen	5.29e-03
2	82	100.0	12	35	W78835	Tetanus toxoid protei	5.29e-03
3	82	100.0	13	30	W50109	Pan DR binding peptid	5.29e-03
4	82	100.0	452	2	R12471	Tetanus toxin fragmen	5.29e-03
5	82	100.0	573	2	P70345	Portion of B fragment	5.29e-03
6	82	100.0	618	33	W48909	SOD-1/TTC hybrid prot	5.29e-03
7	78	95.1	12	1	R06308	Tetanus toxin epitope	1.79e-02
8	51	62.2	401	25	W33109	Streptomyces sp. lami	4.37e+01
9	51	62.2	548	4	R22197	Sequence encoded by b	4.37e+01
10	50	61.0	522	35	W69726	Saccharomyces cerevis	5.72e+01
11	50	61.0	526	5	R26174	Protein disulphide is	5.72e+01
12	49	59.8	220	24	W12652	Orange pectin methyle	7.49e+01
13	49	59.8	584	37	W72964	Aspergillus aculeatus	7.49e+01
14	49	59.8	584	24	W12651	Orange pectin methyle	7.49e+01
15	49	59.8	1218	19	W05833	Human Serrate-1 (HJ1)	7.49e+01
16	49	59.8	1346	30	W31346	Rat tumour suppressor	7.49e+01

17	49	59.8	1596	30	W31347	Rat tumour suppressor	7.49e+01
18	48	58.5	57	38	W88767	Polyptide fragment	9.77e+01
19	47	57.3	821	25	W26664	Yeast checkpoint cont	1.27e+02
20	46	56.1	208	34	W38500	S. pneumoniae 50S rib	1.65e+02
21	46	56.1	306	15	R75132	Porcine carboxypeptid	1.65e+02
22	46	56.1	329	20	W06174	Mature HCPB with PelB	1.65e+02
23	46	56.1	349	21	W13763	Carboxypeptidase B mu	1.65e+02
24	46	56.1	349	21	W13761	Carboxypeptidase B mu	1.65e+02
25	46	56.1	349	20	W06182	Modified HCPB (D253R)	1.65e+02
26	46	56.1	349	21	W13759	Carboxypeptidase B mu	1.65e+02
27	46	56.1	349	21	W13762	Carboxypeptidase B mu	1.65e+02
28	46	56.1	349	21	W13750	Carboxypeptidase B mu	1.65e+02
29	46	56.1	349	21	W13765	Carboxypeptidase B mu	1.65e+02
30	46	56.1	349	21	W13752	Carboxypeptidase B mu	1.65e+02
31	46	56.1	349	21	W13757	Carboxypeptidase B mu	1.65e+02
32	46	56.1	349	20	W06173	mature HCPB-(His)6-C-	1.65e+02
33	46	56.1	349	21	W13764	Carboxypeptidase B mu	1.65e+02
34	46	56.1	349	21	W13751	Carboxypeptidase B mu	1.65e+02
35	46	56.1	349	21	W13758	Carboxypeptidase B mu	1.65e+02
36	46	56.1	349	21	W13753	Carboxypeptidase B mu	1.65e+02
37	46	56.1	349	21	W13760	Carboxypeptidase B mu	1.65e+02
38	46	56.1	349	20	W06181	Modified HCPB (D253K)	1.65e+02
39	46	56.1	404	15	R75131	Porcine Tyr-His-Met P	1.65e+02
40	46	56.1	415	20	W06172	Human pancreatic carb	1.65e+02
41	46	56.1	415	35	W74476	Human pancreatic carb	1.65e+02
42	46	56.1	424	20	W06175	ProHCPB with PelB lea	1.65e+02
43	46	56.1	613	35	W74481	Human pancreatic carb	1.65e+02
44	46	56.1	655	4	R20139	Sequence encoded by n	1.65e+02
45	46	56.1	715	4	R20138	Sequence encoded by P	1.65e+02

ALIGNMENTS

RESULT 1
ID W67035 standard; peptide; 12 AA.
AC W67035;
DT 15-DEC-1998 (first entry)
DE Tetanus toxin fragment (residues 1273-1284).
KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;
KW dendrimeric poly-lysine; epitope; tumour.
OS Clostridium tetani.
PN W09843677-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; E01922.
PR 27-MAR-1997; US-041726.
PA (INSP) INST PASTEUR.
PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;
DR WPI; 98-557071/47.
PT Carbohydrate peptide conjugate used as vaccine - comprises carrier with dendrimeric poly-lysine enabling multiple epitopes to be covalently attached
PT covalently attached
PS Disclosure; Page 13; 55pp; English.
CC The invention relates to a new carbohydrate peptide conjugate, which comprises a carrier with a dendrimeric poly-lysine enabling multiple epitopes to be covalently attached to it. Also claimed are: (1) an antibody purified from biological fluid or cells of organisms administered with the carbohydrate peptide conjugate, and (2) a diagnosis kit comprising antigen-specific antibodies elicited by immunisation with the carbohydrate peptide conjugate. The peptide conjugate, antibody and diagnosis kit are used to provide pharmaceutical compositions and vaccines against tumours. These can be used to support an immune response against viral infections caused by hepatitis virus, HIV or cytomegalo virus. They can be used to enhance immune responses, especially B- and T-cell responses, of humans and animals against bacterial infections. The carbohydrate peptide conjugate stimulates the antibody and T-cell response without stimulating undesired immune responses. The composition is capable of increasing the survival of tumour bearing humans and animals. The present sequence corresponds to residues 1273-1284 of tetanus toxin. The synthetic peptide corresponding to this sequence may be used as an epitope in a carbohydrate peptide conjugate.
SQ Sequence 12 AA;

Query Match 100.0%; Score 82; DB 36; Length 12;

Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gqigndpnrdil 12
|||||
QY 1 GQIGNDPNRDIL 12

RESULT 2
ID W78835 standard; peptide; 12 AA.
AC W78835;
DT 17-NOV-1998 (first entry)
DE Tetanus toxoid protein fragment 1273-1284.
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;
KW class II associated peptide; pathogen; gene therapy; genetic disease;
KW infection; downregulation; immune response.
OS Synthetic.
PN WO9831398-A1.
PD 23-JUL-1998.
PF 22-JAN-1998; U01499.
PR 06-JAN-1998; US-003253.
PR 22-JAN-1997; US-787547.
PA (PANG-) PANGAEA PHARM INC.
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;
WPI; 98-427556/36.
PT New preparations of microparticles - comprising a synthetic polymer
PT matrix and nucleic acid comprising an expression vector for use in
PT gene therapy
PS Disclosure; Page 8; 101pp; English.
CC A microparticle preparation (MP) has been developed, consisting of
CC microparticles having a diameter of less than 100 mu m. The MP comprises:
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers
CC having a solubility in water of less than 1 mg/l; and (b) an expression
CC vector selected from RNA molecules (at least 50% of which are closed
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)
CC a PM; and (b) a NAM comprising an expression control sequence operatively
CC linked to a coding sequence, where the coding sequence encodes an
CC expression product selected from: (1) a polypeptide at least 7 amino
CC acids in length, having a sequence identical to the sequence of: (i) a
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment
CC of a naturally-occurring protein from an infectious agent which infects
CC a mammal; (2) a peptide having a length and sequence which permits it to
CC bind to an MHC class I or II molecule; and (3) the polypeptide or the
CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793
CC to W78897 are peptide fragments for use in the present invention. The
CC MPs are highly effective vehicles for the delivery of polynucleotides
CC into phagocytic cells. They can be used for gene therapy, e.g. for
CC treating genetic diseases, infections or tumours or for downregulating
CC an immune response.
SQ Sequence 12 AA;

Query Match 100.0%; Score 82; DB 35; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 gqigndpnrdil 12
|||||
QY 1 GQIGNDPNRDIL 12

RESULT 3
ID W50109 standard; peptide; 13 AA.
AC W50109;
DT 30-JUN-1998 (first entry)
DE Pan DR binding peptide (6).
KW Pan DR binding peptide; antigen binding site; MHC molecule;
KW DR locus.
OS Synthetic.
PN US5736142-A.
PD 07-APR-1998.
PF 14-SEP-1994; 305871.
PR 14-SEP-1994; US-305871.

PR 14-SEP-1993; US-121101.
PA (CYTE-) CYTEL CORP.
PI Alexander JL, Gaeta F, Grey HM, Sette A, Sidney J;
WPI; 98-239154/21.
PT Peptides that bind to MHC molecules of all DR alleles - inhibiting
PT or inducing MHC Class II mediated activation of T cells
PS Disclosure; Columns 33-34; 29pp; English.
CC The present sequence, a pan DR binding peptide, is capable of
CC binding antigen binding sites on MHC molecules, which are encoded
CC by most of the alleles of a DR locus. The peptide can be used to
CC inhibit or induce MHC Class II mediated activation of T-cells or
CC helper T-cells, which themselves mediate a CTL response. The
CC peptide can be used in mammals, especially humans, to inhibit
CC T-cell-mediated events involved in allograft rejection, allergic
CC responses and autoimmunity and as a vaccine adjuvant for enhancing
CC an immune response against an administered immunogen. The peptide
CC can be used with other immunogens to treat, e.g. prostate cancer,
CC hepatitis B, hepatitis C, AIDS, renal and cervical carcinoma,
CC lymphoma, CMV and condyloma acuminatum.
SQ Sequence 13 AA;

Query Match 100.0%; Score 82; DB 30; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 gqigndpnrdil 13
|||||
QY 1 GQIGNDPNRDIL 12

RESULT 4
ID R12471 standard; Protein; 452 AA.
AC R12471;
DT 05-AUG-1991 (first entry)
DE Tetanus toxin fragment C encoded by gene with increased G-C content.
KW Terminator; vaccine.
OS Synthetic.
PN EP-430645-A.
PD 05-JUN-1991.
PF 27-NOV-1990; 312870.
PR 28-NOV-1989; GB-026832.
PR 17-MAR-1990; GB-006097.
PA (WELL) WELLCOME FOUNDATION LTD.
PI Makoff AJ, Romanos MA, Clare JJ, Fairweather NF;
WPI; 91-166115/23.
DR N-PSDB; Q12121.
PT DNA sequence encoding tetanus toxin fragment C - useful in the
PT manufacture of vaccines for immunity to tetanus utilising yeast
PT as host organism.
PS Disclosure; Fig 2; 50pp; English.
CC The (G+C) content of the synthetic gene is increased by 47% wrt the
CC native sequence. This eliminates six "terminator" regions which
CC were found to be present in (A+T) rich regions. The terminators
CC (termination/endo-nucleolytic processing/polyadenylation sites)
CC were previously responsible for incomplete transcription of the
CC mRNA. The elimination of these elements (using codon degeneracy)
CC provided for successful expression in yeast of the tetanus toxin
CC fragment C.
SQ Sequence 452 AA;

Query Match 100.0%; Score 82; DB 2; Length 452;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 gqigndpnrdil 421
|||||
QY 1 GQIGNDPNRDIL 12

RESULT 5
ID P70345 standard; Protein; 573 AA.
AC P70345;
DT 22-APR-1991 (first entry)

DE Portion of B fragment and all of the C fragment of tetanus toxin.
KW TT; vaccine.
OS Clostridium tetani.
PN EP-209281-A.
PD 21-JAN-1987.
PF 27-JUN-1986; 305029.
PR 28-JUN-1985; GB-016442.
PA (WELL) Wellcome Foundation Ltd.
PI Fairweather NF;
DR WPI; 87-015999/03.
DR N-PSDB; N70545.
PT Cloned DNA sequence coding for tetanus toxin - or its fragments
PT contg. epitope used to express antigens for vaccine production.
PS Claim 4; Fig 1; 36pp; English.
CC Gene product comprises a tetanus toxin fragment, which may be
CC expressed in a transformed host, and used as an antigen in vaccine
CC production, against the disease.
SQ Sequence 573 AA;

Query Match 100.0%; Score 82; DB 2; Length 573;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 531 gqigndpnrtil 542
QY 1 GQIGNDPNRDIL 12

RESULT 6
ID W48909 standard; Protein; 618 AA.
AC W48909;
DT 23-SEP-1998 (first entry)
DE SOD-1/TTC hybrid protein.
KW Chimeric; copper-zinc superoxide dismutase; SOD-1; TTC; SOD:Tet451;
KW tetanus toxin fragment C; tetanus holotoxin; nerve cell; stroke;
KW neurological disorder; oxidative stress; brain hypoxia-reperfusion;
KW epilepsy; Parkinson's disease; Huntington's disease.
OS Chimeric - Homo sapiens.
OS Chimeric - Clostridium tetani.
FH Key Location/Qualifiers
FT Region 1..163
FT /note= "SOD-1"
FT /note= "TTC moiety"
FT 168..618
PN US5780024-A.
PD 14-JUL-1998.
PF 21-JUN-1996; 668381.
PR 23-JUN-1995; US-000473.
PR 21-JUN-1996; US-668381.
PA (GEHO) GEN HOSPITAL CORP.
PA (UYMA-) UNIV MARYLAND BALTIMORE.
PI Brown RH, Fishman PS, Francis JW, Hosler BA;
DR WPI; 98-412999/35.
DR N-PSDB; V32580.
PT New hybrid protein of superoxide dismutase and tetanus toxin
PT fragment C - having increased uptake by neurons and retention of
PT enzymatic activity in these cells, for treating neurological
PT diseases associated with oxidative stress
PS Claim 7; Columns 23-26; 23pp; English.
CC The present sequence represents an enzymatically active human
CC copper-zinc superoxide dismutase (SOD-1) fused at its carboxyl
CC terminus with the tetanus toxin fragment C (TTC) moiety. The TTC
CC moiety constitutes amino acid residues 865-1315 of the tetanus
CC holotoxin. The hybrid protein, referred as SOD:Tet451, is claimed
CC to have the following properties: (a) it exhibits Cu/Zn SOD enzymatic
CC activity; (b) the TTC moiety selectively binds to nerve cells and
CC allows uptake of the hybrid protein into these cells; and (c) it
CC retains substantial SOD enzymatic activity following cellular uptake.
CC SOD:Tet451 is claimed to be useful for treating neurological disorders
CC associated with oxidative stress, e.g. stroke, brain hypoxia-reperfusion,
CC epilepsy, Parkinson's and Huntington's diseases.
SQ Sequence 618 AA;

Query Match 100.0%; Score 82; DB 33; Length 618;
Best Local Similarity 100.0%; Pred. No. 5.29e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 576 gqigndpnrtil 587
QY 1 GQIGNDPNRDIL 12

RESULT 7
ID R06308 standard; peptide; 12 AA.
AC R06308;
DT 04-DEC-1990 (first entry)
DE Tetanus toxin epitope.
KW Tetanus toxin; vaccine; major histocompatibility complex; MHC;
KW antimalarial.
OS Synthetic.
PN EP-378881-A.
PD 25-JUL-1990.
PF 27-DEC-1989; 203318.
PR 17-JAN-1989; IT-019110.
PR 16-NOV-1989; IT-022409.
PA (ENIE) ENRICERCHE SPA.
PI Pessi A, Bianchi E, Verdini AS, Corradin G;
DR WPI; 90-225582/30.
PT Synthetic peptide(s) corresp. to tetanus toxin epitope(s) - used
PT as universal carriers for prepn. of immunogenic conjugate(s) for
PT use as vaccines.
PS Claim 1; Page 17; 20pp; English.
CC Epitopic peptides may be used with synthetic hapten derived from:
CC a pathogen to generate an immune response to the pathogen.
CC Peptides are recognised by numerous T-helper cell clones within
CC the context of a wide range of alleles of the human MHC.
CC The peptides may be used in an antimalarial vaccine inducing Ab.
CC response to P.falciparum.
SQ Sequence 12 AA;

Query Match 95.1%; Score 78; DB 1; Length 12;
Best Local Similarity 91.7%; Pred. No. 1.79e-02;
Matches 11; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 geigndpnrtil 12
QY 1 GQIGNDPNRDIL 12

RESULT 8
ID W33109 standard; Protein; 401 AA.
AC W33109;
DT 28-JAN-1998 (first entry)
DE Streptomyces sp. laminary pentose forming enzyme.
KW Laminary pentose forming enzyme; recombinant production.
OS Streptomyces sp.
FH Key Location/Qualifiers
FT Peptide 1..35
FT /label= sig_peptide
FT 36..401
FT /label= mat_peptide
PN J09262090-A.
PD 07-OCT-1997.
PF 28-MAR-1996; 074227.
PR 28-MAR-1996; JP-074227.
PA (DNIN) DAINIPPON INK & CHEM INC.
DR WPI; 97-544153/50.
DR N-PSDB; T88370.
PT DNA encoding laminary pentose forming enzyme - useful for large
PT scale recombinant production
PS Claim 1; Pages 7-9; 10pp; Japanese.
CC The DNA encoding the present sequence, Streptomyces sp. DIC-108
CC laminary pentose forming enzyme, is useful for the large scale
CC recombinant production of the enzyme.
SQ Sequence 401 AA;


```
Query Match      62.2%; Score 51; DB 25; Length 401;
Best Local Similarity 72.7%; Pred. No. 4.37e+01;
Matches      8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 139 gnpdnpndil 149
QY 2 QIGNDPNRDIL 12

RESULT 9
ID R22197 standard; Protein; 548 AA.
AC R22197;
DT 24-JUL-1992 (first entry)
DE Sequence encoded by beta-1, 3-glucanase gene.
KW Recombinant enzyme; glucanase; PUV5-G1S.
OS Oerskovia xanthineolytica.
FH Key Location/Qualifiers
FT peptide 1..36
FT /label= signal
FT region 444..448
FT /label= repeat
FT region 456..474
FT /label= repeat
FT region 488..492
FT /label= repeat
FT region 501..519
FT /label= repeat
PN WO9203557-A.
PD 05-MAR-1992.
PF 16-AUG-1991; CA0294.
PR 17-AUG-1990; US-568869.
PA (CANA ) NAT RES COUNCIL CAN.
PI Shen SH, Chretien P, Bastien L, Slilaty SN;
DR WPI; 92-096900/12.
DR N-PSDB; Q22957.
PT Recombinant beta-1,3-glucanase enzyme - produced by transforming
PT E. coli cells with expression vectors and free of protease
PT activity
PS Example; Fig 2; 32pp; English.
CC The inventors claim a recombinant enzyme prepn. comprising beta-1,
CC 3-glucanase or a mutant or variant and a vector contg. DNA encoding
CC it under the control of an exogenous promoter pref. lac UV5. The
CC example uses the native beta-1, 3-glucanase gene isolated from
CC Oerskovia xanthineolytica genomic DNA and expressed in PUV5-G1S.
SQ Sequence 548 AA;

Query Match      62.2%; Score 51; DB 4; Length 548;
Best Local Similarity 63.6%; Pred. No. 4.37e+01;
Matches      7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 138 gndsdpnrnil 148
QY 2 QIGNDPNRDIL 12

RESULT 10
ID W69726 standard; Protein; 522 AA.
AC W69726;
DT 16-NOV-1998 (first entry)
DE Saccharomyces cerevisiae protein disulphide isomerase.
KW Saccharomyces cerevisiae; yeast; protein disulphide isomerase; PDI;
KW recombinant; endoplasmic reticulum; localisation signal;
KW food processing.
OS Saccharomyces cerevisiae.
PN WO9835049-A1.
PD 13-AUG-1998.
PF 06-FEB-1998; J00498.
PR 07-FEB-1997; JP-038588.
PA (ORIY ) ORIENTAL YEAST CO LTD.
PI Ishii N, Matsu Y, Suzuki Y, Tanaka H, Uchida K;
DR WPI; 98-447246/38.
DR N-PSDB; V50470.
PT Active recombinant yeast protein di:sulphide isomerase - lacking
```

```
PT localisation signal, can be produced in quantity for use in food
PT processing
PS Disclosure; Fig 2-4; 48pp; Japanese.
CC The present sequence represents yeast (Saccharomyces cerevisiae)
CC protein disulphide isomerase (PDI). The present invention describes a
CC biologically active recombinant yeast PDI, in which the endoplasmic
CC reticulum localisation signal (His-Asp-Glu-Leu) at the C-terminal is
CC defective or lacking, is produced by: (a) modifying a gene encoding the
CC natural PDI by adding, deleting or substituting one or more bases in the
CC region encoding the endoplasmic reticulum localisation signal so that
CC part of all of the localisation signal fails; (b) inserting this
CC modified gene into a suitable vector (such as the yeast expression
CC vector YEpIG11); (c) transforming a host cell with the vector; (d)
CC culturing the transformant at a pH near to neutrality (preferably at
CC pH 6.5 to 8.0), and (e) isolating the PDI from the culture medium. PDI
CC is used in the processing of foods such as ham, sausages, fish products
CC and bean-curd (tofu). The modified PDI retains its biological activity
CC but is expressed into the culture medium from where it can be isolated
CC by simple purification methods in high yield.
SQ Sequence 522 AA;

Query Match      61.0%; Score 50; DB 35; Length 522;
Best Local Similarity 54.5%; Pred. No. 5.72e+01;
Matches      6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 eivndpkkdvl 398
QY 2 QIGNDPNRDIL 12

RESULT 11
ID R26174 standard; Protein; 526 AA.
AC R26174;
DT 04-FEB-1993 (first entry)
DE Protein disulphide isomerase.
KW yeast; isomerisation; catalyst; enzyme; disulphide bonds.
OS Saccharomyces cerevisiae.
PN J04197176-A.
PD 16-JUL-1992.
PF 28-NOV-1990; 322620.
PR 28-NOV-1990; JP-322620.
PA (TOFU ) TONEN CORP.
DR WPI; 92-288434/35.
DR N-PSDB; Q27388.
PT yeast protein di:sulphide isomerase - used to give active
PT conformation to protein having inadequate di:sulphide bond
PS Disclosure; Fig 15; 22pp; Japanese.
CC The sequence given is a yeast protein disulphide isomerase (PDI)
CC which catalyses the isomerisation of disulphide bonds. It has a
CC molecular weight of approx 70 kD and has an optimum pH for activity
CC of 8.75. This enzyme can be used to give active conformation to a
CC protein having inadequate disulphide bonds. The gene encoding this
CC protein can be used to produce expression vectors for the large scale
CC production of PDI.
SQ Sequence 526 AA;

Query Match      61.0%; Score 50; DB 5; Length 526;
Best Local Similarity 54.5%; Pred. No. 5.72e+01;
Matches      6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Db 388 eivndpkkdvl 398
QY 2 QIGNDPNRDIL 12

RESULT 12
ID W12662 standard; Protein; 220 AA.
AC W12662;
DT 13-NOV-1997 (first entry)
DE Orange pectin methylesterase isoform L extension sequence.
KW Pectin methylesterase; PME; orange; de-esterified pectin;
KW esterification; fruit juice; yoghurt; whey protein; casein;
KW protein stabilisation.
```

OS Citrus sinensis.
 PN W09703574-A1.
 PD 06-FEB-1997.
 PF 12-JUL-1996; E03051.
 PR 14-JUL-1995; GB-014438.
 PA (DANI-) DANISCO AS.
 PI Buchholt HC, Christensen TMIE, Kreiberg JD, Nielsen J;
 PI Rasmussen P, Thorsoe H;
 DR WPI; 97-145190/13.
 DR N-PSDB; T51740.
 PT Stabilising protein in acidic soln. - by adding enzymatically
 PT de-esterified pectin, produced using pectin methylesterase
 PS Claim 44; Page 75; 114pp; English.
 CC This 220-amino acid polypeptide comprises an N-terminal extension
 CC sequence of orange pectin methyltransferase (PME). The extension
 CC is located N-terminally to the signal peptide of the PME (W12661)
 CC encoded by PME cDNA clone p034 (T51739). The 64 kDa longer (L)
 CC isoform of PME is believed to be more heat stable than the 36 kDa
 CC shorter (L) isoform. Isoform S is believed to start the initial
 CC de-esterification of a pectin and is then superseded by isoform L.
 CC The 220-amino acid extension can be used to impart or increasing
 CC heat stability to a protein.
 SQ Sequence 220 AA;
 Query Match 59.8%; Score 49; DB 24; Length 220;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 12 gdnngnephhail 23
 QY 1 GQIGNDPNRDIL 12
 RESULT 13
 ID W72964 standard; Protein; 584 AA.
 AC W72964;
 DT 21-JAN-1999 (first entry)
 DE Aspergillus aculeatus pectin methyl esterase.
 KW Orange; pectin methyl esterase; PME; Aspergillus; foodstuff; marmalade;
 KW fruit; vegetable; jam; preserve.
 OS Aspergillus aculeatus.
 PN W09847391-A1.
 PD 29-OCT-1998.
 PF 24-APR-1998; IB0673.
 PR 24-APR-1997; GB-008278.
 PA (DANI-) DANISCO AS.
 PI Christensen TMIE, Hyttel S, Kreiberg JD;
 DR WPI; 98-583317/49.
 DR N-PSDB; V64074.
 PT New composition containing pectin methyl esterase and two substrates
 PT - induces gelling in foodstuffs having less than 50% soluble solids
 PT content, used for, e.g. preparation of foodstuff from fruits
 PS Disclosure; Page 44; 29pp; English.
 CC A composition has been developed which comprises a pectin methyl
 CC esterase (PME), and two PME substrates either one of which does not
 CC originate in situ from the other. The present sequence represents a PME
 CC from Aspergillus aculeatus. Also described in the present invention are:
 CC (1) a method of preparing the above composition by forming a mixture of
 CC a PME and two PME substrates either one of which does not originate in
 CC situ from the other; (2) a process comprising adding to a PME substrate,
 CC a PME and another PME substrate, where neither PME substrate originates
 CC in situ from the other; (3) a process of imparting stability to a
 CC reaction medium comprising a PME substrate, by adding at least 1 PME and
 CC a second PME substrate, where neither PME substrate originates in situ
 CC from the other; (4) an aqueous system being in a solidified gel state
 CC and having a soluble solids content of less than 50% w/w, where the
 CC gelling has occurred by use of a high ester PME substrate; and (5) a
 CC foodstuff comprising or prepared from or by the invention. The methods
 CC and products are used in the preparation of foodstuffs from fruit or
 CC vegetable materials containing pectin, such as jams and preserves. The
 CC process is advantageous over prior art in that it comprises addition of
 CC a second PME substrate which will overcome problems associated with
 CC differing quantities and qualities of PME substrates.

SQ Sequence 584 AA;
 Query Match 59.8%; Score 49; DB 37; Length 584;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 58 gdnngnephhail 69
 QY 1 GQIGNDPNRDIL 12
 RESULT 14
 ID W12661 standard; Protein; 584 AA.
 AC W12661;
 DT 13-NOV-1997 (first entry)
 DE Orange pectin methylesterase (S isoform).
 DE Pectin methylesterase; PME; orange; de-esterified pectin;
 KW esterification; fruit juice; yoghurt; whey protein; casein;
 KW protein stabilisation.
 OS Citrus sinensis.
 FH Key Location/Qualifiers
 FT peptide 1..46
 FT /label= Sig_peptide
 FT protein 47..584
 FT /label= Mat_protein
 PN W09703574-A1.
 PD 06-FEB-1997.
 PF 12-JUL-1996; E03051.
 PR 14-JUL-1995; GB-014438.
 PA (DANI-) DANISCO AS.
 PI Buchholt HC, Christensen TMIE, Kreiberg JD, Nielsen J;
 PI Rasmussen P, Thorsoe H;
 DR WPI; 97-145190/13.
 DR N-PSDB; T51739.
 PT Stabilising protein in acidic soln. - by adding enzymatically
 PT de-esterified pectin, produced using pectin methylesterase
 PS Claim 18; Page 73; 114pp; English.
 CC A polypeptide comprises a novel orange pectin methylesterase (PME)
 CC that can be obtained by expression of the PME coding sequence
 CC contained in NCIMB 40750 or of a cDNA clone (T51739) obtained from
 CC an orange fruit cDNA library. A longer isoform (see W12662) of the
 CC PME was also identified. Recombinant PME can be expressed in
 CC transformed host cells or transgenic organisms, and used in a
 CC claimed method for de-esterifying a pectin. The de-esterified
 CC pectin is used in claimed methods for stabilising a protein in an
 CC acidic environment (such as fruit juice, drinking yoghurt or drinks
 CC containing whey or milk proteins) without adversely affecting the
 CC viscosity of that environment. The recombinant enzyme can also be
 CC used to esterify pectins in non-aqueous medium. Another novel PME
 CC (see W12660) is also claimed.
 SQ Sequence 584 AA;
 Query Match 59.8%; Score 49; DB 24; Length 584;
 Best Local Similarity 50.0%; Pred. No. 7.49e+01;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 Db 58 gdnngnephhail 69
 QY 1 GQIGNDPNRDIL 12
 RESULT 15
 ID W05833 standard; Protein; 1218 AA.
 AC W05833;
 DT 28-JAN-1997 (first entry)
 DE Human Serrate-1 (HJ1).
 KW Serrate-1; human jagged-1; HJ1; Notch; cell differentiation;
 KW cell fate; central nervous system; cancer; tissue repair; therapy;
 KW diagnosis; antibody.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT domain 1..1067
 FT /label= Extracellular_domain

FT peptide 14..29
FT /label= sig_peptide
FT domain 185..229
FT /label= DSL
FT /note= "region of homology with Drosophila Delta
FT and Serrate, predicted to mediate binding
FT with Notch"
FT 234..896
FT /label= ELR
FT /note= "epidermal growth factor-like repeat domain"
FT region 234..264
FT /label= ELR1
FT region 265..299
FT /label= ELR2
FT region 300..339
FT /label= ELR3
FT region 340..377
FT /label= ELR4
FT region 378..415
FT /label= ELR5
FT region 416..453
FT /label= ELR6
FT region 454..490
FT /label= ELR7
FT region 491..528
FT /label= ELR8
FT region 529..566
FT /label= ELR9
FT region 567..598
FT /label= Partial_ELR
FT region 599..532
FT /label= Partial_ELR
FT region 633..670
FT /label= ELR10
FT region 671..708
FT /label= ELR11
FT region 709..747
FT /label= ELR12
FT region 748..785
FT /label= ELR13
FT region 786..823
FT /label= ELR14
FT region 824..862
FT /label= ELR15
FT region 863..879
FT /label= Partial_ELR
FT region 880..896
FT /label= Partial_ELR
FT domain 1068..1089
FT /label= Transmembrane_domain
FT domain 1090..1218
FT /label= Intracellular_domain

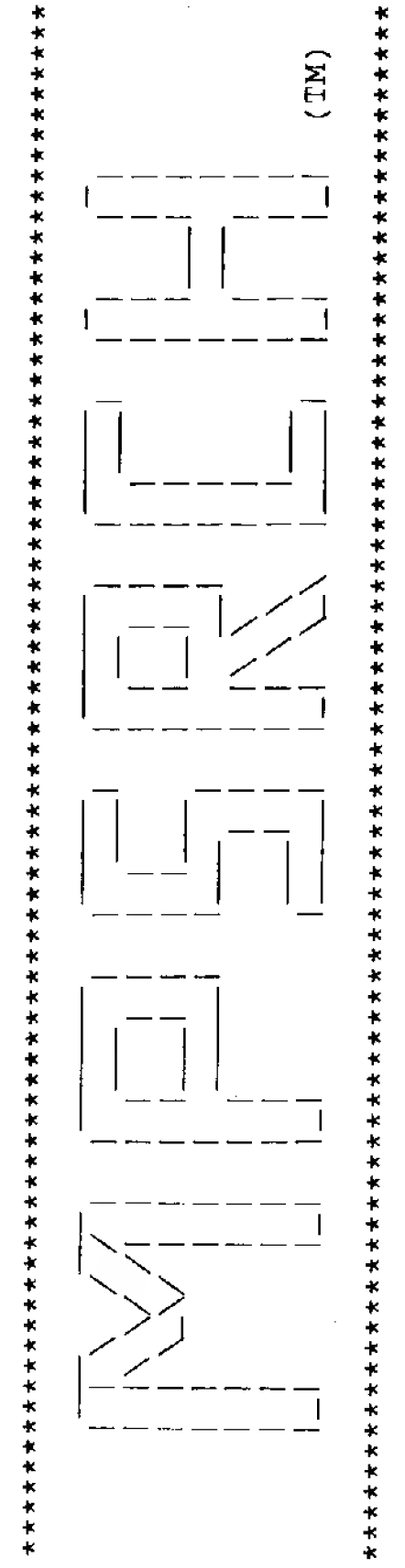
PN WO9627610-A1.
PD 12-SEP-1996.
PF 07-MAR-1996; U03172.
PR 07-MAR-1995; US-400159.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PA (UYA) UNIV YALE.
PI Artavanis-Tsakonas S, Gray GE, Henrique DMP, Ish-Horowicz D;
PI Lewis JH, Mann RS, Myat AM;
DR WPI; 96-425379/42.
DR N-PSDB; T40090.
PT Vertebrate Serrate protein and related DNA - used to treat or
PT prevent malignancies characterised by increased Notch activity.
PS Claim 4; Page 95-98; 161pp; English.
CC Human Serrate-1 (W05833) and human Serrate-2 (W05833) are ligands
CC for the zygotic neurogenic locus Notch, and are believed to play a
CC major role in determining cell fates (differentiation) in the
CC central nervous system. Their amino acid sequences were deduced
CC from cDNA clones (see also T40090-91) isolated from human foetal
CC brain cDNA libraries. The proteins, antibodies raised to them,
CC and encoding nucleic acids can be used in the detection of
CC Serrate sequences and in the treatment of disorders of cell fate

CC or differentiation, partic. cancer, nervous system disorders
CC and in tissue repair or regeneration.
SQ Sequence 1218 AA;

Query Match 59.8%; Score 49; DB 19; Length 1218;
Best Local Similarity 77.8%; Pred. No. 7.49e+01;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Db 114 gndpnrivl 122
| | | | | : |
QY 4 GNDPNRDIL 12

Search completed: Wed Aug 4 15:27:34 1999
Job time : 13 secs.



MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run On: Wed Aug 4 15:29:30 1999; MasPar time 8.88 Seconds
Tabular output not generated. 66.583 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pap
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 403756 seqs, 49297578 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending
1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWU8 20:NEWU9

Statistics: Mean 17.656; Variance 44.251; scale 0.399

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES								
Result No.	Score	Query		Length	DB	ID	Description	Pred. No.
		Match						
1	82	100.0	12	1	PCT-US98-0	Sequence 46, Applicati	5.50e-03	
2	82	100.0	12	15	US-09-049-	Sequence 3, Applicatio	5.50e-03	
3	82	100.0	12	18	US-09-321-	Sequence 46, Applicati	5.50e-03	
4	82	100.0	12	15	US-09-003-	Sequence 46, Applicati	5.50e-03	
5	82	100.0	13	12	US-08-788-	Sequence 7, Applicatio	5.50e-03	
6	82	100.0	13	9	US-08-485-	Sequence 6, Applicatio	5.50e-03	
7	82	100.0	13	9	US-08-485-	Sequence 6, Applicatio	5.50e-03	
8	82	100.0	13	7	US-08-305-	Sequence 6, Applicatio	5.50e-03	
9	82	100.0	14	5	US-08-121-	Sequence 22, Applicati	5.50e-03	
10	82	100.0	14	15	US-09-082-	Sequence 511, Applicat	5.50e-03	
11	82	100.0	1315	14	US-08-913-	Sequence 1, Applicatio	5.50e-03	
12	53	64.6	373	2	US-60-096-	Sequence 19382, Applic	4.69e+01	
13	53	64.6	373	17	US-09-248-	Sequence 19382, Applic	4.69e+01	
14	50	61.0	20	7	US-08-374-	Sequence 20, Applicati	1.11e+02	
15	50	61.0	432	7	US-08-374-	Sequence 11, Applicati	1.11e+02	
16	50	61.0	504	4	US-08-089-	Sequence 18, Applicati	1.11e+02	
17	50	61.0	521	10	US-08-569-	Sequence 14, Applicati	1.11e+02	
18	50	61.0	521	17	US-09-262-	Sequence 32, Applicati	1.11e+02	
19	50	61.0	522	7	US-08-310-	Sequence 1, Applicatio	1.11e+02	
20	50	61.0	530	10	US-08-569-	Sequence 17, Applicati	1.11e+02	
21	50	61.0	530	17	US-09-262-	Sequence 35, Applicati	1.11e+02	

22	50	61.0	635	17	US-09-216- Sequence 2, Applicatio	1.11e+02
23	50	61.0	3052	10	US-08-569- Sequence 8, Applicatio	1.11e+02
24	50	61.0	3052	17	US-09-262- Sequence 26, Applicati	1.11e+02
25	49	59.8	164	2	US-60-096- Sequence 17431, Applic	1.48e+02
26	49	59.8	164	17	US-09-248- Sequence 17431, Applic	1.48e+02
27	49	59.8	175	2	US-60-142- Sequence 424, Applicat	1.48e+02
28	49	59.8	175	2	US-60-141- Sequence 438, Applicat	1.48e+02
29	49	59.8	190	2	US-60-140- Sequence 722, Applicat	1.48e+02
30	49	59.8	220	14	US-08-983- Sequence 5, Applicatio	1.48e+02
31	49	59.8	584	14	US-08-983- Sequence 2, Applicatio	1.48e+02
32	49	59.8	841	15	US-09-092- Sequence 10, Applicati	1.48e+02
33	49	59.8	1208	16	US-09-199- Sequence 1, Applicatio	1.48e+02
34	49	59.8	1218	11	US-08-611- Sequence 6, Applicatio	1.48e+02
35	49	59.8	1218	16	US-09-195- Sequence 6, Applicatio	1.48e+02
36	49	59.8	1346	1	PCT-US97-0 Sequence 2, Applicatio	1.48e+02
37	49	59.8	1346	11	US-08-665- Sequence 2, Applicatio	1.48e+02
38	49	59.8	1346	11	US-08-635- Sequence 2, Applicatio	1.48e+02
39	49	59.8	1346	14	US-08-978- Sequence 2, Applicatio	1.48e+02
40	49	59.8	1596	14	US-08-978- Sequence 4, Applicatio	1.48e+02
41	49	59.8	1596	1	PCT-US97-0 Sequence 4, Applicatio	1.48e+02
42	49	59.8	1596	11	US-08-665- Sequence 4, Applicatio	1.48e+02
43	48	58.5	57	1	PCT-US98-1 Sequence 467, Applicat	1.96e+02
44	48	58.5	57	17	US-09-205- Sequence 545, Applicat	1.96e+02
45	48	58.5	186	16	US-09-134- Sequence 5111, Applica	1.96e+02

ALIGNMENTS

RESULT 1

ID PCT-US98-01499-46 STANDARD; PRI; 12 AA.

XX AC xxxxxx

XX DT

XX DE

XX Sequence 46, Application PC/TUS9801499

CC Sequence 46, Application PC/TUS9801499

CC GENERAL INFORMATION:

CC APPLICANT: Pangaea, Inc.

CC TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY

CC TITLE OF INVENTION: OF NUCLEIC ACID

CC NUMBER OF SEQUENCES: 107

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Fish & Richardson, P.C.

CC STREET: 225 Franklin Street

CC CITY: Boston

CC STATE: MA

CC COUNTRY: US

CC ZIP: 02110-2804

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: Windows95

CC SOFTWARE: FastSeq for Windows Version 2.0

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: PCT/US98/01499

CC FILING DATE: 22-JAN-1998

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/787,547

CC FILING DATE: 22-JAN-1997

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Fraser, Janis K.

CC REGISTRATION NUMBER: 34,819

CC REFERENCE/DOCKET NUMBER: 08191/003WO1

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 617-542-5070

CC TELEFAX: 617-542-8906

CC TELEX: 200154

CC INFORMATION FOR SEQ ID NO: 46:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 12 amino acids

CC TYPE: amino acid

```
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
      |||||
QY      1 GQIGNDPNRDIL 12

RESULT 2
ID      US-09-049-847-3      STANDARD;      PRI;      12 AA.
XX      xxxxxx
XX
DT
XX
DE      Sequence 3, Application US/09049847
XX
CC      Sequence 3, Application US/09049847
CC      GENERAL INFORMATION:
CC      APPLICANT: Bay, Sylvie
CC      APPLICANT: Cantacuzene, Daniele
CC      APPLICANT: Leclerc, Claude
CC      APPLICANT: Lo-Man, Richard
CC      TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine
CC      TITLE OF INVENTION: comprising the same and use thereof
CC      FILE REFERENCE: 102.166A
CC      CURRENT APPLICATION NUMBER: US/09/049,847
CC      CURRENT FILING DATE: 1998-03-27
CC      EARLIER APPLICATION NUMBER: 60/041,726
CC      EARLIER FILING DATE: 1997-03-27
CC      NUMBER OF SEQ ID NOS: 4
CC      SOFTWARE: PatentIn Ver. 2.0
CC      SEQ ID NO 3
CC      LENGTH: 12
CC      TYPE: PRT
CC      ORGANISM: Clostridium tetani
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
      |||||
QY      1 GQIGNDPNRDIL 12

RESULT 3
ID      US-09-321-346-46      STANDARD;      PRT;      12 AA.
XX      xxxxxx
XX
DT
XX
DE      Sequence 46, Application US/09321346
XX
CC      Sequence 46, Application US/09321346
CC      GENERAL INFORMATION:
CC      APPLICANT: Lunsford, Lynn B.
CC      APPLICANT: Putnam, David
CC      APPLICANT: Hedley, Mary Lynn
CC      TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC      FILE REFERENCE: 08191/014001
CC      CURRENT APPLICATION NUMBER: US/09/321,346
CC      CURRENT FILING DATE: 1999-05-27
CC      EARLIER APPLICATION NUMBER: US 09/266,463
CC      EARLIER FILING DATE: 1999-03-11
CC      NUMBER OF SEQ ID NOS: 114
```

```
CC      SOFTWARE: FastSEQ for Windows Version 3.0
CC      SEQ ID NO 46
CC      LENGTH: 12
CC      TYPE: PRT
CC      ORGANISM: Clostridium tetani
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 18; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
      |||||
QY      1 GQIGNDPNRDIL 12

RESULT 4
ID      US-09-003-253-46      STANDARD;      PRT;      12 AA.
XX      xxxxxx
XX
DT
XX
DE      Sequence 46, Application US/09003253
XX
CC      Sequence 46, Application US/09003253
CC      GENERAL INFORMATION:
CC      APPLICANT: Hedley, Mary Lynne
CC      APPLICANT: Curley, Joanne M.
CC      APPLICANT: Langer, Robert S.
CC      TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY OF NUCLEIC ACID
CC      NUMBER OF SEQUENCES: 108
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Fish & Richardson P.C.
CC      STREET: 225 Franklin Street
CC      CITY: Boston
CC      STATE: MA
CC      COUNTRY: USA
CC      ZIP: 02110-2804
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Diskette
CC      COMPUTER: IBM Compatible
CC      OPERATING SYSTEM: Windows 95
CC      SOFTWARE: FastSEQ for Windows Version 2.0b
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: US/09/003,253
CC      FILING DATE: 06-JAN-1998
CC      CLASSIFICATION: 514
CC      PRIOR APPLICATION DATA:
CC      APPLICATION NUMBER: 60/035,983
CC      FILING DATE: 22-JAN-1997
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Fraser, Ph.D., J.D., Janis K.
CC      REGISTRATION NUMBER: 34,819
CC      REFERENCE/DOCKET NUMBER: 08191/003002
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: 617/542-5070
CC      TELEFAX: 617/542-8906
CC      TELEX: 200154
CC      INFORMATION FOR SEQ ID NO: 46:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 12 amino acids
CC      TYPE: amino acid
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: peptide
SQ      SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match      100.0%; Score 82; DB 15; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db      1 GQIGNDPNRDIL 12
      |||||
```


QY 1 GQIGNDPNRDIL 12

RESULT 5
ID US-08-788-822A-7 STANDARD; PRT; 13 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX

Sequence 7, Application US/08788822A

Sequence 7, Application US/08788822A
GENERAL INFORMATION:
APPLICANT: Alexander, Jeffrey L.
APPLICANT: Defrees, Shawn
APPLICANT: Sette, Alessandro
TITLE OF INVENTION: Induction of Immune Response Against
TITLE OF INVENTION: Desired Determinants
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/788,822A
FILING DATE: 23-JAN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/010,510
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bastian, Kevin L.
REGISTRATION NUMBER: 34,774
REFERENCE/DOCKET NUMBER: 014137-009210JS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 13 AA; 1426 MW; 637 CN;

Query Match 100.0%; Score 82; DB 12; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GQIGNDPNRDIL 13
| | | | | | | | | |
QY 1 GQIGNDPNRDIL 12

RESULT 6
ID US-08-485-218-6 STANDARD; PRT; 13 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX

Sequence 6, Application US/08485218

CC Sequence 6, Application US/08485218
CC GENERAL INFORMATION:
CC APPLICANT: Sette, Alessandro
CC APPLICANT: Gaeta, Federico
CC APPLICANT: Grey, Howard M
CC APPLICANT: Sidney, John
CC APPLICANT: Alexander, Jeffery L
CC APPLICANT: Del Guercio, Marie-France
CC TITLE OF INVENTION: Alteration of Immune Response Using Pan
CC TITLE OF INVENTION: DR-Binding Peptides
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend and Crew
CC STREET: One Market Plaza, Steuart Street Tower
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94105-1492
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/485,218
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/305,871
CC FILING DATE: 14-SEP-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 14137-62-2
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-543-9600
CC TELEFAX: 415-543-5043
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC SEQUENCE 13 AA; 1426 MW; 537 CN;

Query Match 100.0%; Score 82; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GQIGNDPNRDIL 13
| | | | | | | | | |
QY 1 GQIGNDPNRDIL 12

RESULT 7
ID US-08-485-218A-6 STANDARD; PRT; 13 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
CC
CC

Sequence 6, Application US/08485218A

Sequence 6, Application US/08485218A
GENERAL INFORMATION:
APPLICANT: Sette, Alessandro
APPLICANT: Gaeta, Federico

```
CC APPLICANT: Grey, Howard M.
CC APPLICANT: Sidney, John
CC APPLICANT: Alexander, Jeffery L.
CC APPLICANT: Del Guercio, Marie-France
CC TITLE OF INVENTION: Alteration of Immune Response Using Pan
CC TITLE OF INVENTION: DR-Binding Peptides
CC NUMBER OF SEQUENCES: 31
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Crew LLP
CC STREET: Two Embarcadero Center, Eighth Floor
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-3834
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/485,218A
CC FILING DATE: 07-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/305,871
CC FILING DATE: 14-SEP-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 018623-006220US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 576-0200
CC TELEFAX: (415) 576-0300
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1426 MW; 637 CN;

Query Match 100.0%; Score 82; DB 9; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GQIGNDPNRDIL 13
   | | | | | | | | | |
QY 1 GQIGNDPNRDIL 12

RESULT 8
ID US-08-305-871-6 STANDARD; PRI; 13 AA.
XX
AC xxxxxxx
XX
DT
XX
DE
XX
Sequence 6, Application US/08305871
GENERAL INFORMATION:
APPLICANT: Sette, Alessandro
APPLICANT: Gaeta, Federico
APPLICANT: Grey, Howard M.
APPLICANT: Sidney, John
APPLICANT: Alexander, Jeffrey L.
TITLE OF INVENTION: Alteration of Immune Response Using Pan
TITLE OF INVENTION: DR-Binding peptides
NUMBER OF SEQUENCES: 17
```

```
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Crew LLP
CC STREET: Two Embarcadero Center, Eighth Floor
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94111-3834
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/305,871
CC FILING DATE: 14-SEP-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 14137-62-1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 576-0200
CC TELEFAX: (415) 576-0300
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1426 MW; 637 CN;

Query Match 100.0%; Score 82; DB 7; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.50e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GQIGNDPNRDIL 13
   | | | | | | | | | |
QY 1 GQIGNDPNRDIL 12

RESULT 9
ID US-08-121-101-22 STANDARD; PRI; 14 AA.
XX
AC xxxxxxx
XX
DT
XX
DE
XX
Sequence 22, Application US/08121101
GENERAL INFORMATION:
APPLICANT: SETTE, Alessandro
APPLICANT: GAETA, Federico
APPLICANT: GREY, Howard M.
APPLICANT: SIDNEY, John
TITLE OF INVENTION: ALTERATION OF IMMUNE RESPONSE USING PAN
TITLE OF INVENTION: DR-BINDING PEPTIDES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
```

CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/121,101
 CC FILING DATE: 14-SEP-1993
 CC CLASSIFICATION: 514
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Bastian, Kevin L.
 CC REGISTRATION NUMBER: 34,774
 CC REFERENCE/DOCKET NUMBER: 14137-62
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (415) 543-9600
 CC TELEFAX: (415) 543-5043
 CC INFORMATION FOR SEQ ID NO: 22:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 14 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: unknown
 CC MOLECULE TYPE: peptide
 CC HYPOTHETICAL: NO
 CC SEQUENCE 14 AA; 1589 MW; 743 CN;

Query Match 100.0%; Score 82; DB 5; Length 14;
 Best Local Similarity 100.0%; Pred. No. 5.50e-03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 3 GOIGNDPNRDIL 14
 |||||
 QY 1 GOIGNDPNRDIL 12

RESULT 10
 ID US-09-082-279-511 STANDARD; PRT; 14 AA.

XX xxxxxx

Sequence 511, Application US/09082279

Sequence 511, Application US/09082279

CC GENERAL INFORMATION:
 CC APPLICANT: Barney, Shawn
 CC APPLICANT: Guthrie, Kelly
 CC APPLICANT: Merutka, Gene
 CC APPLICANT: Anwer, Mohamed
 CC APPLICANT: Lambert, Dennis
 CC TITLE OF INVENTION: HYBRID POLYPEPTIDES WITH ENHANCED
 CC TITLE OF INVENTION: PHARMACOKINETIC PROPERTIES
 CC FILE REFERENCE: 7853-043
 CC CURRENT APPLICATION NUMBER: US/09/082,279
 CC CURRENT FILING DATE: 1998-05-20
 CC NUMBER OF SEQ ID NOS: 1436
 CC SOFTWARE: FastSeq for Windows Version 3.0
 CC SEQ ID NO 511
 CC LENGTH: 14
 CC TYPE: PRT
 CC ORGANISM: Artificial Sequence
 CC FEATURE:
 CC OTHER INFORMATION: Core polypeptide
 CC SEQUENCE 14 AA; 1589 MW; 903 CN;

Query Match 100.0%; Score 82; DB 15; Length 14;
 Best Local Similarity 100.0%; Pred. No. 5.50e-03;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GOIGNDPNRDIL 13
 |||||
 QY 1 GOIGNDPNRDIL 12

RESULT 11
 ID US-08-913-880-1 STANDARD; PRT; 1315 AA.

XX

AC xxxxxx

XX

DT

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

SQ

Sequence 1, Application US/08913880

Sequence 1, Application US/08913880

GENERAL INFORMATION:

APPLICANT: MATSUDA, Morihiro

TITLE OF INVENTION: TETANUS TOXIN FUNCTIONAL FRAGMENT ANTIIGEN AND TETANUS

TITLE OF INVENTION: VACCINE

FILE REFERENCE: 216-380P

CURRENT APPLICATION NUMBER: US/08/913,880

CURRENT FILING DATE: 1997-09-24

NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 1315

TYPE: PRT

ORGANISM: Clostridium tetani

SEQUENCE 1315 AA; 150681 MW; 9110516 CN;

Query Match 100.0%; Score 82; DB 14; Length 1315;

Best Local Similarity 100.0%; Pred. No. 5.50e-03;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1273 GOIGNDPNRDIL 1284

|||||

QY 1 GOIGNDPNRDIL 12

RESULT 12

ID US-60-096-409-19382 STANDARD; PRT; 373 AA.

XX xxxxxx

XX

DT

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 64.6%; Score 53; DB 2; Length 373;

Best Local Similarity 58.3%; Pred. No. 4.69e+01;

Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 183 SQIGIDPPRGVL 194

:||| ||| :|

QY 1 GOIGNDPNRDIL 12

RESULT 13

ID US-09-248-796-19382 STANDARD; PRT; 373 AA.

XX

AC xxxxxx

XX

DT

XX

DE

Sequence 19382, Application US/09248796

```
xx Sequence 19382, Application US/09248796
cc GENERAL INFORMATION:
cc APPLICANT: Keith Weinstock et al
cc TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALB
cc TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
cc FILE REFERENCE: 107196.132
cc CURRENT APPLICATION NUMBER: US/09/248,796
cc CURRENT FILING DATE: 1999-02-12
cc NUMBER OF SEQ ID NOS: 28206
cc SEQ ID NO 19382
cc LENGTH: 373
cc TYPE: PRT
cc ORGANISM: Candida albicans
cc SEQUENCE 373 AA; 41473 MW; 672994 CN;

Query Match 64.6%; Score 53; DB 17; Length 373;
Best Local Similarity 58.3%; Pred. No. 4.69e+01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 183 SQIGDPPRGVL 194
QY :||| ||| :|
1 GOIGNDPNRDIL 12

RESULT 14
ID US-08-374-888A-20 STANDARD; PRT; 20 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 20, Application US/08374888A
Sequence 20, Application US/08374888A
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
APPLICANT: Ren, Dejian
APPLICANT: Zheng, Wei
APPLICANT: Dubald, Manuel Marcel Paul
TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, LLP
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,888A
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm M.
REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 022650-263
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 20 AA; 1952 MW; 1883 CN;
```

```
Query Match 61.0%; Score 50; DB 7; Length 20;
Best Local Similarity 50.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

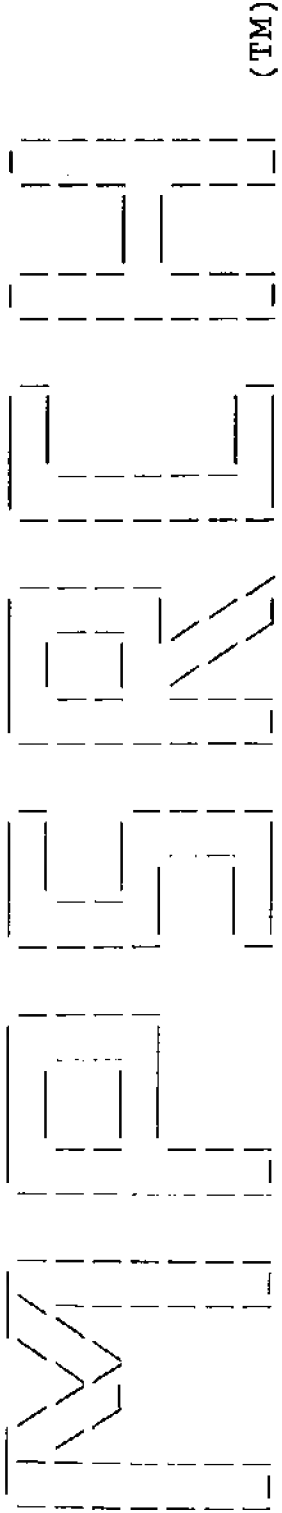
Db 2 GOAGAEPSRDSM 13
QY ||| :||| :
1 GOIGNDPNRDIL 12

RESULT 15
ID US-08-374-888-11 STANDARD; PRT; 432 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 11, Application US/08374888
Sequence 11, Application US/08374888
GENERAL INFORMATION:
APPLICANT: HALL, Linda M
APPLICANT: REN, Dejian
APPLICANT: ZHENG, Wei
APPLICANT: DUBALD, Manuel Marcel Paul
TITLE OF INVENTION: GENES ENCODING AN INSECT CALCIUM CHANNEL
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,888
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K
REGISTRATION NUMBER: P39,300
REFERENCE/DOCKET NUMBER: 022650-263
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE 432 AA; 47886 MW; 945243 CN;

Query Match 61.0%; Score 50; DB 7; Length 432;
Best Local Similarity 50.0%; Pred. No. 1.11e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 160 GOAGAEPSRDSM 171
QY ||| :||| :
1 GOIGNDPNRDIL 12
```

Search completed: Wed Aug 4 15:29:44 1999
Job time : 14 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:29:07 1999; MasPar time 2.30 seconds
Tabular output not generated. 52.893 Million cell updates/sec

Title: >US-09-049-847-3
Description: (1-12) from US09049847.pep
Perfect Score: 82
Sequence: 1 GQIGNDPNRDIL 12

Scoring table: PAM 150
Gap 15

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued

1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 16.008; Variance 42.182; scale 0.379

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	82	100.0	12	2	US-08-787-	Sequence 46, Applicati	1.05e-03
2	82	100.0	13	1	US-08-305-	Sequence 6, Applicatio	1.05e-03
3	82	100.0	452	1	US-08-280-	Sequence 4, Applicatio	1.05e-03
4	82	100.0	452	1	US-07-618-	Sequence 2, Applicatio	1.05e-03
5	82	100.0	452	1	US-07-618-	Sequence 4, Applicatio	1.05e-03
6	82	100.0	452	1	US-08-110-	Sequence 8, Applicatio	1.05e-03
7	82	100.0	452	1	US-08-280-	Sequence 2, Applicatio	1.05e-03
8	82	100.0	618	2	US-08-668-	Sequence 5, Applicatio	1.05e-03
9	50	61.0	504	2	US-08-441-	Sequence 18, Applicati	1.80e-01
10	50	61.0	521	2	US-08-557-	Sequence 32, Applicati	1.80e-01
11	50	61.0	530	2	US-08-557-	Sequence 35, Applicati	1.80e-01
12	50	61.0	3052	2	US-08-557-	Sequence 26, Applicati	1.80e-01
13	49	59.8	1218	2	US-08-400-	Sequence 6, Applicatio	2.39e-01
14	47	57.3	776	2	US-08-870-	Sequence 17, Applicati	4.19e-01
15	47	57.3	776	1	US-08-198-	Sequence 17, Applicati	4.19e-01
16	47	57.3	821	1	US-08-198-	Sequence 6, Applicatio	4.19e-01
17	47	57.3	821	1	US-08-870-	Sequence 4, Applicatio	4.19e-01
18	46	56.1	306	1	US-08-696-	Sequence 6, Applicatio	5.53e-01
19	46	56.1	404	1	US-08-696-	Sequence 2, Applicatio	5.53e-01
20	46	56.1	655	2	US-08-469-	Sequence 27, Applicati	5.53e-01
21	46	56.1	655	2	US-08-469-	Sequence 28, Applicati	5.53e-01
22	45	54.9	30	1	US-08-145-	Sequence 19, Applicati	7.28e-01
23	45	54.9	3666	2	US-08-222-	Sequence 12, Applicati	7.28e-01

24	45	54.9	3727	2	US-08-222-	Sequence 27, Applicati	7.28e-01
25	45	54.9	3778	2	US-08-222-	Sequence 2, Applicatio	7.28e-01
26	43	52.4	153	4	5196194-20	Patent No. 5196194.	1.25e+02
27	43	52.4	708	1	US-08-396-	Sequence 8, Applicatio	1.25e-02
28	43	52.4	708	1	US-08-818-	Sequence 8, Applicatio	1.25e-02
29	43	52.4	739	1	US-08-818-	Sequence 10, Applicati	1.25e-02
30	43	52.4	739	1	US-08-396-	Sequence 10, Applicati	1.25e-02
31	43	52.4	1068	1	US-08-818-	Sequence 12, Applicati	1.25e-02
32	43	52.4	1068	1	US-08-396-	Sequence 12, Applicati	1.25e-02
33	43	52.4	1075	3	PCT-US94-0	Sequence 41, Applicati	1.25e-02
34	42	51.2	15	4	5183734-5	Patent No. 5183734.	1.64e+02
35	42	51.2	16	4	5183734-13	Patent No. 5183734.	1.64e+02
36	42	51.2	16	4	5183734-12	Patent No. 5183734.	1.64e+02
37	42	51.2	112	2	US-08-961-	Sequence 7, Applicatio	1.64e+02
38	42	51.2	112	2	US-08-283-	Sequence 7, Applicatio	1.64e+02
39	42	51.2	112	2	US-08-283-	Sequence 2, Applicatio	1.64e+02
40	42	51.2	112	1	US-08-326-	Sequence 17, Applicati	1.64e+02
41	42	51.2	153	4	5204446-4	Patent No. 5204446.	1.64e+02
42	42	51.2	154	4	5183734-1	Patent No. 5183734.	1.64e+02
43	42	51.2	202	3	PCT-US96-1	Sequence 16, Applicati	1.64e+02
44	42	51.2	568	3	PCT-US94-0	Sequence 30, Applicati	1.64e+02
45	42	51.2	1528	1	US-08-326-	Sequence 2, Applicatio	1.64e+02

ALIGNMENTS

RESULT 1
ID US-08-787-547-46 STANDARD; PRT; 12 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
XX

Sequence 46, Application US/08787547

Sequence 46, Application US/08787547

Patent No. 5783567

GENERAL INFORMATION:

APPLICANT: Hedley, Mary Lynne

APPLICANT: Curley, Joanne M.

APPLICANT: Langer, Robert S.

TITLE OF INVENTION: MICROPARTICLES FOR DELIVERY

TITLE OF INVENTION: OF NUCLEIC ACID

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/787,547

FILING DATE: 22-JAN-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 08191/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 12 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 12 AA; 1311 MW; 550 CN;

Query Match 100.0%; Score 82; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GQIGNDPNRDIL 12
|
QY 1 GQIGNDPNRDIL 12

RESULT 2
ID US-08-305-871A-6 STANDARD; PRT; 13 AA.

AC xxxxxx
XX
DT
XX
DE Sequence 6, Application US/08305871A
XX Sequence 6, Application US/08305871A
CC Patent No. 5736142
CC GENERAL INFORMATION:
CC APPLICANT: Sette, Alessandro
CC APPLICANT: Gaeta, Federico
CC APPLICANT: Grey, Howard M.
CC APPLICANT: Sidney, John
CC APPLICANT: Alexander, Jeffrey L.
CC TITLE OF INVENTION: Alteration of Immune Response Using Pan
CC TITLE OF INVENTION: DR-Binding Peptides
CC NUMBER OF SEQUENCES: 29
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Townsend and Townsend and Crew LLP
CC STREET: Two Embarcadero Center, Eighth Floor
CC CITY: San Francisco
CC STATE: California
CC COUNTRY: USA

CC ZIP: 94111-3834
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/305,871A
CC FILING DATE: 14-SEP-1994
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/121,101
CC FILING DATE: 14-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 14137-0062-10
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 576-0200
CC TELEFAX: (415) 576-0300
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 13 AA; 1426 MW; 637 CN;

Query Match 100.0%; Score 82; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2 GQIGNDPNRDIL 13
|
QY 1 GQIGNDPNRDIL 12

RESULT 3
ID US-08-280-228-4 STANDARD; PRT; 452 AA.

AC xxxxxx
XX
DT
XX
DE Sequence 4, Application US/08280228
XX Sequence 4, Application US/08280228
CC Patent No. 5571694
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 452 AA; 51770 MW; 1065961 CN;

Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 GQIGNDPNRDIL 421
|
QY 1 GQIGNDPNRDIL 12

```

XX Sequence 4, Application US/07618312A
DE Sequence 4, Application US/07618312A
XX Patent No. 5389540
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: 14th Floor
CC STREET: 2200 Clarendon Boulevard,
CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/618,312A
CC FILING DATE: 19910516
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Crawford Mr, Arthur R
CC REGISTRATION NUMBER: 25,327
CC REFERENCE/DOCKET NUMBER: 510-51
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 0101 703 8750400
CC TELEFAX: 0101 703 5253468
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51770 MW; 1065961 CN;
SQ
Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12
|||||

RESULT 6
ID US-08-110-786A-8 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
XX
XX
DE Sequence 8, Application US/08110786A
XX
CC Sequence 8, Application US/08110786A
CC Patent No. 5443966
CC GENERAL INFORMATION:
CC APPLICANT: FAIRWEATHER, Neil Fraser
CC APPLICANT: MAKOFF, Andrew Joseph

```

CC TITLE OF INVENTION: Expression of tetanus toxin fragment C
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Nixon & Vanderhye P.C.
CC STREET: 1100 No. 5443965th Glebe Road
CC CITY: Arlington
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/110,786A
CC FILING DATE: 23-AUG-1993 1991
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/777,337
CC FILING DATE: 29-NOV-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/GB90/00943
CC FILING DATE: 20-JUN-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8914122.0
CC FILING DATE: 20 June 1989
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Mary J. Wilson
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-134
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51784 MW; 1060967 CN;
SQ
Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12
RESULT 7
ID US-08-280-228-2 STANDARD; PRT; 452 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08280228
XX Sequence 2, Application US/08280228
CC Patent No. 5571594
CC GENERAL INFORMATION:
CC APPLICANT: Makoff Dr, Andrew J
CC APPLICANT: Romanos Dr, Michael A
CC APPLICANT: Clare Dr, Jeffrey J
CC APPLICANT: Fairweather Dr, Neil F
CC TITLE OF INVENTION: VACCINES
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 No. 5571694th Glebe Road

CC CITY: Arlington,
CC STATE: Virginia
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/280,228
CC FILING DATE: 25-JUL-1994
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/618,312
CC FILING DATE: 27-NOV-1990
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 8926832.0
CC FILING DATE: 28-NOV-1989
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: GB 9006097.1
CC FILING DATE: 17-MAR-1990
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Wilson, Mary J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 117-163
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000
CC TELEFAX: (703) 816-4100
CC TELEX: 200797 NIXN UR
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 452 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 452 AA; 51797 MW; 1065526 CN;
SQ
Query Match 100.0%; Score 82; DB 1; Length 452;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 410 GQIGNDPNRDIL 421
QY 1 GQIGNDPNRDIL 12
RESULT 8
ID US-08-668-381A-5 STANDARD; PRT; 618 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 5, Application US/08668381A
XX Sequence 5, Application US/08668381A
CC Patent No. 5780024
CC GENERAL INFORMATION:
CC APPLICANT: Brown, Robert H.
CC APPLICANT: Fishman, Paul S.
CC APPLICANT: Francis, Jonathan W.
CC APPLICANT: Hosler, Betsy A.
CC TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN
CC NUMBER OF INVENTION: FRAGMENT C HYBRID PROTEIN
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson P.C.
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA

CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/668,381A
CC FILING DATE: 21-JUN-1996
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/000,473
CC FILING DATE: 23-JUN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Clark, Paul T.
CC REGISTRATION NUMBER: 30,164
CC REFERENCE/DOCKET NUMBER: 00786/269001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617/542-5070
CC TELEFAX: 617/542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 618 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 618 AA; 68895 MW; 1991829 CN;
SQ
Query Match 100.0%; Score 82; DB 2; Length 618;
Best Local Similarity 100.0%; Pred. No. 1.05e-03;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 576 QGIGNDPNRDIL 587
QY 1 QGIGNDPNRDIL 12
RESULT 9
ID US-08-441-139-18 STANDARD; PRT; 504 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 18, Application US/08441139
XX
CC Sequence 18, Application US/08441139
CC Patent No. 5773245
CC GENERAL INFORMATION:
CC APPLICANT: Wittrup, Dr. Karl D.
CC APPLICANT: Robinson, Anne S.
CC TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
CC TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
CC NUMBER OF SEQUENCES: 20
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
CC STREET: 400 Garden City Plaza
CC CITY: Garden City
CC STATE: NY
CC COUNTRY: USA
CC ZIP: 11530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,139
CC FILING DATE: 15-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/089,997

CC FILING DATE: 06-JUL-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Digiglio, Frank S.
CC REGISTRATION NUMBER: 31,346
CC REFERENCE/DOCKET NUMBER: 8646
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 516-742-4343
CC TELEFAX: 516-742-4366
CC TELEX: 230 901 SANS UR
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 504 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 504 AA; 55937 MW; 1256202 CN;
SQ
Query Match 61.0%; Score 50; DB 2; Length 504;
Best Local Similarity 54.5%; Pred. No. 1.80e+01;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 371 EIVNDPKKDVIL 381
QY 2 QIGNDPNRDIL 12
RESULT 10
ID US-08-557-122A-32 STANDARD; PRT; 521 AA.
XX
AC xxxxxx
XX
DI
XX
DE Sequence 32, Application US/08557122A
XX
CC Sequence 32, Application US/08557122A
CC Patent No. 5879664
CC GENERAL INFORMATION:
CC APPLICANT: Hjort, Carsten Mailand
CC TITLE OF INVENTION: Fungal Protein Disulfide Isomerase
CC NUMBER OF SEQUENCES: 38
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: No. 5879664o No. 5879664disk of No. 5879664th America, Inc.
CC STREET: 405 Lexington Avenue, 64th Floor
CC CITY: New York
CC STATE: New York
CC COUNTRY: United States of America
CC ZIP: 10174-6401
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/557,122A
CC FILING DATE: 11-DEC-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Lambiris, Elias J.
CC REGISTRATION NUMBER: 33,728
CC REFERENCE/DOCKET NUMBER: 3980.204-US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-867-0123
CC TELEFAX: 212-878-9655
CC INFORMATION FOR SEQ ID NO: 32:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 521 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 521 AA; 58129 MW; 1344734 CN;
SQ

CC STREET: 1155 Avenue of the Americas
 CC CITY: New York
 CC STATE: New York
 CC COUNTRY: USA
 CC ZIP: 10036-2711
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.30
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/400,159
 CC FILING DATE: 07-MAR-1995
 CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Mirock, S. Leslie
 CC REGISTRATION NUMBER: 18,872
 CC REFERENCE/DOCKET NUMBER: 7326-029
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (212) 790-9090
 CC TELEFAX: (212) 869-9741/8864
 CC TELEX: 66141 PENNIE
 CC INFORMATION FOR SEQ ID NO: 6:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 1218 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 1218 AA; 133810 MW; 7064054 CN;
 CC
 CC Query Match 59.8%; Score 49; DB 2; Length 1218;
 CC Best Local Similarity 77.8%; Pred. No. 2.39e+01;
 CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 Db 114 GNDPNRIVL 122
 QY 4 GNDPNRDIL 12
 RESULT 14
 ID US-08-870-693-17 STANDARD; PRT; 776 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 17, Application US/08870693
 XX Sequence 17, Application US/08870693
 CC Patent No. 5866338
 CC GENERAL INFORMATION:
 CC APPLICANT: Hartwell, Leland H.
 CC APPLICANT: Weinert, Ted A.
 CC APPLICANT: Plon, Sharon E.
 CC APPLICANT: Groudine, Mark T.
 CC TITLE OF INVENTION: Cell Cycle Checkpoint Genes
 CC NUMBER OF SEQUENCES: 19
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
 CC STREET: 1420 Fifth Ave., Suite 2800
 CC CITY: Seattle
 CC STATE: WA
 CC COUNTRY: USA
 CC ZIP: 98101-2347
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/870,693
 CC FILING DATE:
 CC CLASSIFICATION: 536

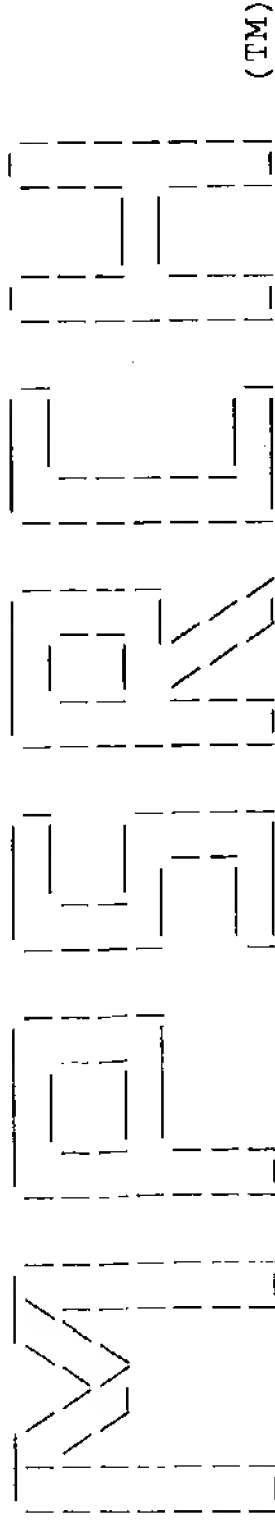
CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/198,446
 CC FILING DATE: February 18, 1994
 CC APPLICATION NUMBER: PCT/US93/04458
 CC FILING DATE: May 12, 1993
 CC APPLICATION NUMBER: US 07/884,426
 CC FILING DATE: May 14, 1992
 CC APPLICATION NUMBER: US 07/882,051
 CC FILING DATE: May 12, 1992
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Sheiness, Diana K.
 CC REGISTRATION NUMBER: 35,356
 CC REFERENCE/DOCKET NUMBER: FHCRI10798
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 206-682-8100
 CC TELEFAX: 206-224-0779
 CC INFORMATION FOR SEQ ID NO: 17:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 776 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC DESCRIPTION: Yeast MEC2 protein
 CC SEQUENCE 776 AA; 87015 MW; 3041193 CN;
 CC
 CC Query Match 57.3%; Score 47; DB 2; Length 776;
 CC Best Local Similarity 45.5%; Pred. No. 4.19e+01;
 CC Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
 CC
 Db 289 GAVGEDAGREI 299
 QY 1 GQIGNDPNRDI 11
 RESULT 15
 ID US-08-198-446B-17 STANDARD; PRT; 776 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 17, Application US/08198446B
 XX Sequence 17, Application US/08198446B
 CC Patent No. 5674996
 CC GENERAL INFORMATION:
 CC APPLICANT: Hartwell, Leland H.
 CC APPLICANT: Weinert, Ted A.
 CC APPLICANT: Plon, Sharon E.
 CC APPLICANT: Groudine, Mark T.
 CC TITLE OF INVENTION: Cell Cycle Checkpoint Genes
 CC NUMBER OF SEQUENCES: 19
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
 CC STREET: 1420 Fifth Ave., Suite 2800
 CC CITY: Seattle
 CC STATE: WA
 CC COUNTRY: USA
 CC ZIP: 98101-2347
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: PatentIn Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/198,446B
 CC FILING DATE: 18-FEB-1994
 CC CLASSIFICATION: 536
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Sheiness, Diana K.
 CC REGISTRATION NUMBER: 35,356
 CC REFERENCE/DOCKET NUMBER: FHCRI17537
 CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 206-682-8100
CC TELEFAX: 206-224-0779
CC INFORMATION FOR SEQ ID NO: 17:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 776 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC DESCRIPTION: yeast MEC2 protein
SQ SEQUENCE 776 AA; 87015 MW; 3041193 CN;

Query Match 57.3%; Score 47; DB 1; Length 776;
Best Local Similarity 45.5%; Pred.No. 4.19e+01;
Matches 5; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Db 289 GAVGEDAGREI 299
|:|:|:|:
QY 1 GQIGNDPNRDI 11

Search completed: Wed Aug 4 15:29:12 1999
Job time : 5 secs.



(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:30:34 1999; MasPar time 4.72 seconds
110.469 Million cell updates/sec
Tabular output not generated.

Title: >US-09-049-847-4
Description: (1-13) from US09049847.ppep
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 122810 seqs, 40068593 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 26.787; Variance 42.449; scale 0.631

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	87	83.7	2207	1	GNNY1P genome polypeptide (v	7.13e-05
2	87	83.7	2209	1	GNNY2P genome polypeptide (v	7.13e-05
3	85	81.7	2205	1	GNNY2W genome polypeptide -	1.75e-04
4	85	81.7	2207	1	GNNY5P genome polypeptide -	1.75e-04
5	85	81.7	2209	1	GNNY3P genome polypeptide -	1.75e-04
6	82	78.8	2207	2	S09553 genome polypeptide -	6.68e-04
7	79	76.0	2206	1	GNNY4P genome polypeptide -	2.49e-03
8	79	76.0	2206	1	S03822 genome polypeptide -	2.49e-03
9	77	74.0	2206	1	GNNY27 genome polypeptide -	5.92e-03
10	70	67.3	1040	2	A57638 receptor tyrosine kin	1.13e-01
11	67	64.4	250	2	D69182 conserved hypotetica	3.82e-01
12	64	61.5	2206	1	GNNY21 genome polypeptide -	1.25e+00
13	62	59.6	874	2	H64228 DNA polymerase III al	2.73e+00
14	61	58.7	149	2	S13460 hemoglobin - southern	4.00e+00
15	61	58.7	255	2	F64503 hypothetical protein	4.00e+00
16	61	58.7	940	2	T01854 hypothetical protein	4.00e+00
17	59	56.7	345	2	F71261 probable lipase - syp	8.48e+00
18	59	56.7	741	2	S39082 myosin heavy chain, e	8.48e+00
19	59	56.7	858	2	S39081 myosin heavy chain, a	8.48e+00
20	59	56.7	955	2	S24348 myosin heavy chain, e	8.48e+00
21	59	56.7	1938	1	JX0178 myosin heavy chain, f	8.48e+00
22	58	55.8	22	2	B20923 light meromyosin - ch	1.23e+01
23	58	55.8	847	2	A64675 alanine--trna ligase	1.23e+01

24	58	55.8	847	2	F71842	alanyl-trna synthetas	1.23e+01
25	58	55.8	874	4	GNHUEP	retrovirus-related po	1.23e+01
26	58	55.8	878	2	A69584	alanine--trna ligase	1.23e+01
27	58	55.8	1332	2	F69732	PBSX prophage ORF xkd	1.77e+01
28	57	54.8	241	2	S62922	probable membrane pro	1.77e+01
29	57	54.8	398	1	R3BYM1	ribosomal protein var	1.77e+01
30	57	54.8	415	2	S29345	translation elongatio	1.77e+01
31	57	54.8	467	2	I49609	proto-oncogene protei	1.77e+01
32	57	54.8	555	2	T00778	tRNA adenylyltransfer	1.77e+01
33	57	54.8	599	2	S65180	hypothetical protein	1.77e+01
34	57	54.8	654	1	P4XRBV	minor inner core prot	1.77e+01
35	57	54.8	936	2	S39083	myosin heavy chain, n	1.77e+01
36	57	54.8	1053	2	D71466	probable ribonucleosi	1.77e+01
37	57	54.8	1940	2	A29320	myosin heavy chain, e	1.77e+01
38	56	53.8	208	2	S76531	hypothetical protein	2.54e+01
39	56	53.8	346	2	C65105	hypothetical 37.3 kD	2.54e+01
40	56	53.8	467	2	A48713	serine/threonine-spec	2.54e+01
41	56	53.8	468	2	H69500	heme biosynthesis pro	2.54e+01
42	56	53.8	1048	2	S27763	Ca2+-transporting ATP	2.54e+01
43	56	53.8	1934	2	I48153	beta-myosin heavy cha	2.54e+01
44	56	53.8	1938	1	S05005	myosin alpha heavy ch	2.54e+01
45	56	53.8	1939	2	I48175	alpha-cardiac myosin	2.54e+01

ALIGNMENTS

RESULT 1
ENTRY genome polypeptide (version 1) - human poliovirus - (strain Mahoney)
TITLE
CONTAINS coat protein VP1; coat protein VP2; coat protein VP3; coat protein VP4; core protein P2-3b; core protein P2-5b; core protein P2-X; genome-linked protein VPg; probable proteinase P3-7c; protein P3-1b; protein P3-2; RNA-directed RNA polymerase (EC 2.7.7.48) P3-4b
ORGANISM #formal_name human poliovirus 1
DATE 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 05-Jun-1998
ACCESSIONS A93258; A90800; A03897
REFERENCE A93258
#authors Kitamura, N.; Semler, B.L.; Rothberg, P.G.; Larsen, G.R.; Adler, C.J.; Dorner, A.J.; Emini, E.A.; Hanecak, R.; Lee, J.J.; van der Werf, S.; Anderson, C.W.; Wimmer, E.
#journal Nature (1981) 291:547-553
#title Primary structure, gene organization and polypeptide expression of poliovirus RNA.
#cross-references MUID:81220953
#accession A93258
#molecule_type genomic RNA
#residues 1-2207 #label K1T1
#cross-references GB:V01148; NID:961236; PID:961237
#note the amino acid sequence of VPg (residues 1543-1564) was also determined and agrees with that shown

REFERENCE A90800
#authors Kitamura, N.; Adler, C.J.; Rothberg, P.G.; Martinko, J.; Nathenson, S.G.; Wimmer, E.
#journal Cell (1980) 21:295-302
#title The genome-linked protein of picornaviruses. VII. Genetic mapping of poliovirus VPg by protein and RNA sequence studies.
#cross-references MUID:81001866
#accession A90800
#molecule_type genomic RNA
#residues 1539-1574 #label K1T2
#note the amino end of VPg corresponds to residue 1543; a choice between the two potential carboxyl cleavage sites, after residue 1564 or 1569, could not be made; the partial sequence of this protein obtained by radiochemical microsequence analysis agrees with that predicted by the virion RNA
REFERENCE A30637
#authors Rothberg, P.G.; Harris, I.J.; Nomoto, A.; Wimmer, E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1978) 75:4868-4872

```
#title
O4-(5'-Uridyl)tyrosine is the bond between the
genome-linked protein and the RNA of poliovirus.
#contents
annotation; chemical characterization
COMMENT VPg is linked by Tyr-1545 to the uridylate residue at the 5' end of
the genome RNA. It is required to initiate RNA synthesis and it
may also be involved in morphogenesis.
COMMENT Coat proteins VP2 and VP3 and the RNA-directed RNA polymerase are
related to their counterparts in foot-and-mouth disease virus.
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; nucleotidyltransferase;
phosphoprotein; polypeptide
FEATURE
2-69
70-340
341-578
579-880
881-1455
1030-1455
#product coat protein VP4 #status predicted #label VP4\
#product coat protein VP2 #status predicted #label VP2\
#product coat protein VP3 #status predicted #label VP3\
#product coat protein VP1 #status predicted #label VP1\
#product core protein P2-3b #status predicted #label
P23\
#product core protein P2-5b #status predicted #label
P25\
#product core protein P2-X #status predicted #label P2X\
#product protein P3-1b #status predicted #label P31\
#product genome-linked protein VPg #status predicted
#label VPg\
#product protein P3-2 #status predicted #label P32\
#product probable proteinase P3-7c #status predicted
#label P37\
#product RNA-directed RNA polymerase P3-4b #status
predicted #label P34\
#binding_site phosphoryl-RNA (Tyr) (covalent) #status
experimental
#length 2207 #molecular-weight 246540 #checksum 7480
SUMMARY
Query Match 83.7%; Score 87; DB 1; Length 2207;
Best Local Similarity 84.6%; Pred. No. 7.13e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 681 KLFVAVWKITYKDT 593
|::|::|::|::|::|
QY 1 KFLAVWKITYKDT 13
RESULT 2
ENTRY genome polypeptide (version 2) - human poliovirus 1 (strain
TITLE Mahoney)
ORGANISM #formal_name human poliovirus 1
DATE 18-Dec-1981 #sequence_revision 18-Dec-1981 #text_change
ACCESSION A03898
REFERENCE A03898
#authors Racaniello, V.R.; Baltimore, D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1981) 78:4887-4891
#title Molecular cloning of poliovirus cDNA and determination of the
complete nucleotide sequence of the viral genome.
#cross-references MUID:82060159
#accession A03898
##molecule_type genomic RNA
##residues 1-2209 ##label RAC
##cross-references GB:J02281; NID:g332886; PID:g332887
##note the authors translated the codon TTC for residue 464 as
Ser
REFERENCE A03636
#authors Ambros, V.; Baltimore, D.
#journal J. Biol. Chem. (1978) 253:5263-5266
#title Protein is linked to the 5' end of poliovirus RNA by a
phosphodiester linkage to tyrosine.
#contents phosphodiester chemical characterization
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS genome-linked protein; nucleotidyltransferase;
phosphoprotein; polypeptide
#product coat protein VP4 #status predicted #label VP4\
```

```
70-341
342-579
580-881
882-1456
1031-1456
#product coat protein P2-5b #status predicted #label
P25\
#product core protein P2-X #status predicted #label P2X\
#product protein P3-1b #status predicted #label P31\
#product genome-linked protein VPg #status predicted
#label VPg\
#product protein P3-2 #status predicted #label P32\
#product probable proteinase P3-7c #status predicted
#label P37\
#product RNA-directed RNA polymerase P3-4b #status
predicted #label P34\
#binding_site phosphoryl-RNA (Tyr) (covalent) #status
experimental
#length 2209 #molecular-weight 246538 #checksum 7583
SUMMARY
Query Match 83.7%; Score 87; DB 1; Length 2209;
Best Local Similarity 84.6%; Pred. No. 7.13e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 682 KLFVAVWKITYKDT 694
|::|::|::|::|::|
QY 1 KFLAVWKITYKDT 13
RESULT 3
ENTRY genome polypeptide - human poliovirus 2 (strain W-2)
TITLE coat protein 1A; coat protein 1B; coat protein 1C; coat
CONTAINS protein 1D; genome-linked protein VPg; nonstructural
protein 2B; nonstructural protein 2C; nonstructural protein
3A; proteinase 2A; proteinase 3C; RNA-directed RNA
polymerase (EC 2.7.7.48)
ORGANISM #formal_name human poliovirus 2
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
ACCESSION A34032
REFERENCE A34032
#authors Pevear, D.C.; Oh, C.K.; Cunningham, L.L.; Calenoff, M.;
Jubelt, B.
#journal J. Gen. Virol. (1990) 71:43-52
#title Localization of genomic regions specific for the attenuated,
mouse-adapted poliovirus type 2 strain W-2.
#cross-references MUID:90155230
#accession A34032
##molecule_type genomic RNA
##residues 1-2205 ##label PEV
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS coat protein; genome-linked protein; nonstructural protein;
nucleotidyltransferase; phosphoprotein; polypeptide;
proteinase
FEATURE
1-69
70-340
341-578
579-879
880-1028
1029-1125
1126-1454
1455-1541
1542-1563
1564-1746
1747-2205
1544
#product coat protein 1A #status predicted #label VP4\
#product coat protein 1B #status predicted #label VP2\
#product coat protein 1C #status predicted #label VP3\
#product coat protein 1D #status predicted #label VP1\
#product proteinase 2A #status predicted #label P2A\
#product nonstructural protein 2B #status predicted
#label P2B\
#product nonstructural protein 2C #status predicted
#label P2C\
#product nonstructural protein 3A #status predicted
#label P3A\
#product genome-linked protein VPg #status predicted
#label VPg\
#product proteinase 3C #status predicted #label P3C\
#product RNA-directed RNA polymerase #status predicted
#label RRP\
#binding_site phosphoryl-RNA (Tyr) (covalent) #status
```

```

SUMMARY          predicted
                  #length 2205 #molecular-weight 245701 #checksum 6218
Query Match      81.7%; Score 85; DB 1; Length 2205;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches          10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWVKITYKDT 693
QY 1 KFLAVWVKITYKDT 13

RESULT 4
ENTRY genome polyprotein - human poliovirus 2 (strain Lansing)
TITLE coat protein VP1; coat protein VP2; coat protein VP3; coat
CONTAINS protein VP4; genome-linked protein VPg; nonstructural
          protein 2B; nonstructural protein 2C; nonstructural protein
          3A; proteinase (EC 3.4.-.-) 2A; proteinase (EC 3.4.-.-) 3C;
          RNA-directed RNA polymerase (EC 2.7.7.48) 3D
          #formal_name human poliovirus 2
ORGANISM #formal_name human poliovirus 2
DATE      31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change
          05-Jun-1998
ACCESSIONS A29507
REFERENCE  A29507
#authors   La Monica, N.; Meriam, C.; Racaniello, V.R.
#journal   J. Virol. (1986) 57:515-525
#title     Mapping of sequences required for mouse neurovirulence of
          poliovirus type 2 Lansing.
#cross-references MUID:86115399
#accession  A29507
#molecule_type genomic RNA
#residues   1-2207 #label LAM
#cross-references GB:M12197; NID:g332890; PID:g332891
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS        coat protein; genome-linked protein; hydrolase; nonstructural
          protein; nucleotidyltransferase; phosphoprotein;
          polyprotein; proteinase

FEATURE
1-69      #product coat protein VP4 #status predicted #label VP4\
70-340    #product coat protein VP2 #status predicted #label VP2\
341-578   #product coat protein VP3 #status predicted #label VP3\
579-879   #product coat protein VP1 #status predicted #label VP1\
880-1028  #product proteinase 2A #status predicted #label P2A\
1029-1125 #product nonstructural protein 2B #status predicted
          #label N2B\
1126-1454 #product nonstructural protein 2C #status predicted
          #label N2C\
1455-1541 #product nonstructural protein 3A #status predicted
          #label N3A\
1542-1563 #product genome-linked protein VPg #status predicted
          #label VPg\
1564-1746 #product proteinase 3C #status predicted #label P3C\
1747-2207 #product RNA-directed RNA polymerase #status predicted
          #label RRP\
1544      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted
SUMMARY    #length 2207 #molecular-weight 245829 #checksum 666

Query Match      81.7%; Score 85; DB 1; Length 2207;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches          10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLFVWVKITYKDI 693
QY 1 KFLAVWVKITYKDT 13

RESULT 5
ENTRY genome polyprotein - human poliovirus 1 (strain Sabin)
TITLE #formal_name human poliovirus 1
ORGANISM #formal_name human poliovirus 1
DATE      17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change

22-Jan-1999
A03899
A03899
#authors   Nomoto, A.; Omata, T.; Toyoda, H.; Kuge, S.; Horie, H.;
          Kataoka, Y.; Genba, Y.; Nakano, Y.; Imura, N.
#journal   Proc. Natl. Acad. Sci. U.S.A. (1982) 79:5793-5797
#title     Complete nucleotide sequence of the attenuated poliovirus
          Sabin 1 strain genome.
#cross-references MUID:83299876
#accession  A03899
#molecule_type genomic RNA
#residues   1-2209 #label NOM
#cross-references GB:J02282; GB:J02285; GB:J02286;
          GB:V01150; GB:V01157; PID:g61258
          NID:g61257;
#note      this virus is a live vaccine strain derived from the
          Mahoney strain by spontaneous mutations during the
          attenuation process
          #superfamily poliovirus genome polyprotein
          genome-linked protein; phosphoprotein; polyprotein
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS        genome-linked protein; phosphoprotein; polyprotein
FEATURE
2-69      #product coat protein VP4 #status predicted #label VP4\
70-341    #product coat protein VP2 #status predicted #label VP2\
342-579   #product coat protein VP3 #status predicted #label VP3\
580-881   #product coat protein VP1 #status predicted #label VP1\
882-1456  #product core protein P2-3b #status predicted #label
          P23\
1031-1456 #product core protein P2-5b #status predicted #label
          P25\
1128-1456 #product core protein P2-X #status predicted #label P2X\
1457-2209 #product protein P3-1b #status predicted #label P31\
1544-1565 #product genome-linked protein VPg #status predicted
          #label VPg\
1566-2209 #product protein P3-2 #status predicted #label P32\
1566-1748 #product probable proteinase P3-7c #status predicted
          #label P37\
1749-2209 #product RNA-directed RNA polymerase P3-4b #status
          predicted #label P34\
1546      #binding_site phosphoryl-RNA (Tyr) (covalent) #status
          predicted
SUMMARY    #length 2209 #molecular-weight 246576 #checksum 5033

Query Match      81.7%; Score 85; DB 1; Length 2209;
Best Local Similarity 76.9%; Pred. No. 1.75e-04;
Matches          10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 682 KLFVWVKITYKDT 694
QY 1 KFLAVWVKITYKDT 13

RESULT 6
ENTRY genome polyprotein - human poliovirus 2 (strain P712)
TITLE coat protein VP1; coat protein VP2; coat protein VP3; coat
CONTAINS protein VP4; core protein P2-3b; core protein P2-5b; core
          protein P2-X; genome-linked protein VPg; protein P3-1b;
          protein P3-2; protein P3-4a; protein P3-6a; protein P3-6b;
          protein P3-9; proteinase P3-7c; RNA-directed RNA polymerase
          (EC 2.7.7.48) P3-4b
          #formal_name human poliovirus 2
ORGANISM #formal_name human poliovirus 2
DATE      07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
          05-Jun-1998
ACCESSIONS S09553
REFERENCE  S03822
#authors   Toyoda, H.; Kohara, M.; Kataoka, Y.; Suganuma, T.; Omata, T.;
          Imura, N.; Nomoto, A.
#journal   J. Mol. Biol. (1984) 174:561-585
#title     Complete nucleotide sequences of all three poliovirus
          serotype genomes. Implication for genetic relationship,
          gene function and antigenic determinants.
          #cross-references MUID:84216300
          #accession  S09553
          #molecule_type genomic RNA
```



```

##residues      1-2207 ##label TOY
##cross-references EMBL:X00595; NID:g61127; PID:e275415; PID:gl628430
CLASSIFICATION  #superfamily poliovirus genome polypeptide
KEYWORDS        genome-linked protein; nucleotidyltransferase;
                phosphoprotein; polypeptide
FEATURE
  1544          #binding_site phosphoryl-RNA (Tyr) (covalent) #status
            predicted
SUMMARY        #length 2207 #molecular-weight 245967 #checksum 1659

Query Match      78.8%; Score 82; DB 2; Length 2207;
Best Local Similarity 69.2%; Pred. No. 6.68e-04;
Matches          9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db      681 RLFSVWKITYKDT 593
      ::::|::|::|::|::|::|
QY      1 KFLAVWKITYKDT 13

RESULT      7
ENTRY       genome polypeptide - human poliovirus 3 (strain Sabin vaccine
TITLE       P3/Leon/37, P3/Leon/12a[1]b)
CONTAINS    coat protein VP1; coat protein VP2; coat protein VP3; coat
            protein VP4; core protein P2-3b; core protein P2-5b; core
            protein P2-X; genome-linked protein VPg; probable
            proteinase P3-7c; protein P3-1b; protein P3-2; RNA-directed
            RNA polymerase (EC 2.7.7.48) P3-4b
ORGANISM    #formal_name human poliovirus 3
DATE        19-Feb-1984 #sequence_revision 03-Aug-1984 #text_change
            26-Feb-1999
ACCESSIONS  A93987; A93484; S42524; A03900
REFERENCE   A93987
#authors    Stanway, G.; Hughes, P.J.; Mountford, R.C.; Reeve, P.; Minor,
            P.D.; Schild, G.C.; Almond, J.W.
#journal    Proc. Natl. Acad. Sci. U.S.A. (1984) 81:1539-1543
#title      Comparison of the complete nucleotide sequences of the
            genomes of the neurovirulent poliovirus P3/Leon/37 and its
            attenuated Sabin vaccine derivative P3/Leon 12a-1b.
#cross-references MUID:84170338
#accession  A93987
##molecule_type genomic RNA
##residues   1-2206 ##label ST1
##cross-references GB:K01392; NID:g332895; PID:g332896
##experimental_source strain Sabin vaccine P3/Leon/37
##note       the strain Sabin vaccine P3/Leon/37 is the progenitor of
            the strain Sabin vaccine P3/Leon 12a[1]b
##note       the authors translated the codon GAU for residue 497 as
            Gly
REFERENCE   A93484
#authors    Stanway, G.; Cann, A.J.; Hauptmann, R.; Hughes, P.; Clarke,
            L.D.; Mountford, R.C.; Minor, P.D.; Schild, G.C.; Almond,
            J.W.
#journal    Nucleic Acids Res. (1983) 11:5629-5643
#title      The nucleotide sequence of poliovirus type 3 lon 12 a-1b:
            comparison with poliovirus type 1.
#cross-references MUID:83299239
#accession  A93484
##molecule_type genomic RNA
##residues   1-430,'P',432-863,'R',865-907,'A',909-2206 ##label ST2
##experimental_source strain Sabin vaccine P3/Leon 12a[1]b
##note       the authors translated the codon GAU for residue 497 as
            Gly
REFERENCE   S42524
#authors    Stanway, G.; Cann, A.J.; Hauptmann, R.; Mountford, R.C.;
            Clarke, L.D.; Reeve, P.; Minor, P.D.; Schild, G.C.; Almond,
            J.W.
#journal    Eur. J. Biochem. (1983) 135:529-533
#title      Nucleic acid sequence of the region of the genome encoding
            capsid protein VP1 of neurovirulent and attenuated type 3
            polioviruses.
#cross-references MUID:84004370
#accession  S42524

```

```

##molecule_type genomic RNA
##residues      579-878 ##label STA
##cross-references EMBL:V01540; NID:g61153; PID:g929811
##experimental_source strain Sabin vaccine P3/Leon/37
COMMENT        Coat proteins VP2 and VP3 and the RNA-directed RNA polymerase are
            related to their counterparts in foot-and-mouth disease virus.
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS        genome-linked protein; nucleotidyltransferase; phosphoprotein
FEATURE
  2-69          #product coat protein VP4 #status predicted #label VP4\
  70-340         #product coat protein VP2 #status predicted #label VP2\
  341-578        #product coat protein VP3 #status predicted #label VP3\
  579-878        #product coat protein VP1 #status predicted #label VP1\
  879-1453       #product core protein P2-3b #status predicted #label
            P23\
  1027-1453      #product core protein P2-5b #status predicted #label
            P25\
  1124-1453      #product core protein P2-X #status predicted #label P2X\
  1454-2206      #product protein P3-1b #status predicted #label P31\
  1541-1562      #product genome-linked protein VPg #status predicted
            #label VPg\
  1563-2206      #product protein P3-2 #status predicted #label P32\
  1563-1745      #product probable proteinase P3-7c #status predicted
            #label P37\
  1746-2206      #product RNA-directed RNA polymerase P3-4b #status
            predicted #label P34\
  1543           #binding_site phosphoryl-RNA (Tyr) (covalent) #status
            predicted
SUMMARY        #length 2206 #molecular-weight 246163 #checksum 7111

Query Match      76.0%; Score 79; DB 1; Length 2206;
Best Local Similarity 69.2%; Pred. No. 2.49e-03;
Matches          9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db      679 KLFAMWRITYKDT 691
      |::|::|::|::|::|
QY      1 KFLAVWKITYKDT 13

RESULT      8
ENTRY       genome polypeptide - human poliovirus 3 (strain Leon 12a1b)
TITLE       coat protein VP1; coat protein VP2; coat protein VP3; coat
            protein VP4; core protein P2-3b; core protein P2-5b; core
            protein P2-X; genome-linked protein VPg; protein P3-1b;
            protein P3-2; protein P3-4a; protein P3-6a; protein P3-6b;
            protein P3-9; proteinase P3-7c; RNA-directed RNA polymerase
            (EC 2.7.7.48) P3-4b
ORGANISM    #formal_name human poliovirus 3
DATE        07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
            05-Jun-1998
ACCESSIONS  S03822; S04920
REFERENCE   S03822
#authors    Toyoda, H.; Kohara, M.; Kataoka, Y.; Suganuma, T.; Omata, T.;
            Imura, N.; Nomoto, A.
#journal    J. Mol. Biol. (1984) 174:561-585
#title      Complete nucleotide sequences of all three poliovirus
            serotype genomes. Implication for genetic relationship,
            gene function and antigenic determinants.
#cross-references MUID:84216300
#accession  S03822
##molecule_type genomic RNA
##residues      1-2206 ##label TOY
##cross-references EMBL:X00596
REFERENCE   S04920
#authors    Nomoto, A.
#submissio  submitted to the EMBL Data Library, April 1985
#accession  S04920
##molecule_type genomic RNA
##residues      1-541,'Q',543-696,'M',698-1311,'E',1313-2206 ##label NOM
##cross-references EMBL:X00596; NID:g61139; PID:g61140
CLASSIFICATION #superfamily poliovirus genome polypeptide
KEYWORDS        genome-linked protein; nucleotidyltransferase;

```



```
Db 187 ELSVWDVPYRIT 198
||:|:|:|:|:|
QY 2 FLAVWKITYKDI 13

RESULT 12
ENTRY GNNY21 #type complete
TITLE genome polyprotein - coxsackievirus A21 (strain Coe)
CONTAINS coat protein 1A; coat protein 1B; coat protein 1C; coat
protein 1D; core protein 2A; core protein 2B; core protein
2C; genome-linked protein VPg; protein 3A; proteinase;
RNA-directed RNA polymerase (EC 2.7.7.48)
ORGANISM #formal_name coxsackievirus A21
DATE 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change
05-Jun-1998
ACCESSIONS A33373
REFERENCE A33373
#authors Hughes, P.J.; North, C.; Minor, P.D.; Stanway, G.
#journal J. Gen. Virol. (1989) 70:2943-2952
#title The complete nucleotide sequence of coxsackievirus A21.
#cross-references MUID:90063544
#accession A33373
##molecule_type genomic RNA
##residues 1-2206 #label HUG
##cross-references GB:D00538; NID:g221147; PID:d1000880; PID:g221148
CLASSIFICATION #superfamily poliovirus genome polyprotein
KEYWORDS coat protein; core protein; genome-linked protein;
nucleotidyltransferase; phosphoprotein; polyprotein;
proteinase
FEATURE
1-69 #product coat protein 1A #status predicted #label VP4\
70-341 #product coat protein 1B #status predicted #label VP2\
342-578 #product coat protein 1C #status predicted #label VP3\
579-881 #product coat protein 1D #status predicted #label VP1\
882-1028 #product core protein 2A #status predicted #label PA2\
1029-1125 #product core protein 2B #status predicted #label PB2\
1126-1453 #product core protein 2C #status predicted #label PC2\
1454-1540 #product protein 3A #status predicted #label PA3\
1541-1562 #product genome-linked protein VPg #status predicted
#label PB3\
1563-1745 #product proteinase #status predicted #label PC3\
1746-2206 #product RNA-directed RNA polymerase #status predicted
#label PD3\
1543 #binding_site phosphoryl-RNA (Tyr) (covalent) #status
predicted
SUMMARY #length 2206 #molecular-weight 246049 #checksum 3669
Query Match 61.5%; Score 64; DB 1; Length 2206;
Best Local Similarity 53.8%; Pred. No. 1.25e+00;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Db 679 KHFNIWNITYTDT 691
||:|:|:|:|:|
QY 1 KFLAVWKITYKDT 13

RESULT 13
ENTRY H64228 #type complete
TITLE DNA polymerase III alpha chain (dnaE) homolog - Mycoplasma
genitalium (SGC3)
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change
21-Aug-1998
ACCESSIONS H64228
REFERENCE A64200
#authors Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton,
R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.;
Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.;
Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.;
Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,
J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.;
Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III,
```

```
C.A.; Venter, J.C.
#journal Science (1995) 270:397-403
#title The minimal gene complement of Mycoplasma genitalium.
#cross-references MUID:96026346
#accession H64228
##status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-874 #label TIGR
##cross-references GB:U39705; GB:L43967; NID:g1045953; PID:g1045954;
TIGR:MG261
##experimental_source strain G-37
GENETICS
#genetic_code SGC3
CLASSIFICATION #superfamily polydeoxyribonucleotide synthase (NAD+)
SUMMARY #length 874 #molecular-weight 100434 #checksum 6779
Query Match 59.6%; Score 62; DB 2; Length 874;
Best Local Similarity 70.0%; Pred. No. 2.73e+00;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 118 LTTWKSTYKD 127
||:|:|:|:|:|
QY 3 LAVWKITYKD 12

RESULT 14
ENTRY S13460 #type complete
TITLE hemoglobin - southern lamprey
ORGANISM #formal_name Mordacia mordax #common_name southern lamprey
DATE 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change
10-Jul-1998
ACCESSIONS S13460
REFERENCE S13458
#authors Hombados, I.; Vidal, Y.; Rodewald, K.; Braunitzer, G.;
Neuzil, E.
#journal Biol. Chem. Hoppe-Seyler (1991) 372:49-56
#title The primary structure of the hemoglobins of a southern
hemisphere lamprey (Mordacia mordax, Cyclostomata).
#cross-references MUID:91248417
#accession S13460
##molecule_type protein
##residues 1-149 #label HOM
FUNCTION
#description in erythrocytes binds and transports molecular oxygen from
lung to tissues
CLASSIFICATION #superfamily globin; globin homology
KEYWORDS chromoprotein; erythrocyte; heme; iron; oxygen carrier
FEATURE
11-149 #domain globin homology #label GLB\
73 #binding_site oxygen (His) (distal axial ligand) #status
predicted\
105 #binding_site heme iron (His) (proximal axial ligand)
#status predicted
SUMMARY #length 149 #molecular-weight 16508 #checksum 7504
Query Match 58.7%; Score 61; DB 2; Length 149;
Best Local Similarity 50.0%; Pred. No. 4.00e-00;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

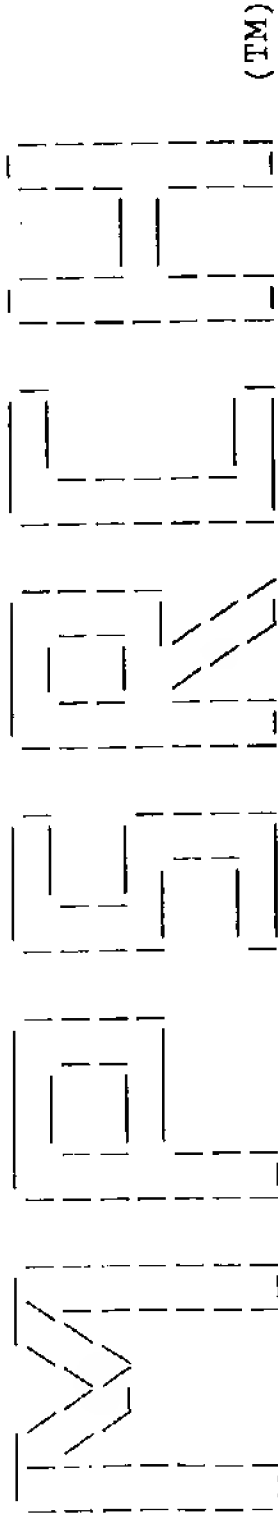
Db 18 KILAAWDLVYKN 29
||:|:|:|:|:|
QY 1 KFLAVWKITYKD 12

RESULT 15
ENTRY F64503 #type complete
TITLE hypothetical protein M1632 - Methanococcus jannaschii
ORGANISM #formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
04-Sep-1998
ACCESSIONS F64503
REFERENCE A64300
```

```
#authors      Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
               R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
               R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
               Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
               Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
               Scott, J.L.; Geoghagen, N.S.M.; Weidman, J.F.; Fuhrmann,
               J.L.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
               J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
               K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
               H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
               Science (1996) 273:1058-1073
#journal      Complete genome sequence of the methanogenic archaeon,
#title        Methanococcus jannaschii.
#cross-references MUID:96337999
#accession    F64503
#status       preliminary; nucleic acid sequence not shown;
               translation not shown
#molecule_type DNA
#residues     1-255 #label BUL
#cross-references GB:U67503; GB:L77117; NID:g1592220; PID:g1592222;
               TIGR:MJ1632; PID:g1511593
GENETICS
#map_position FOR1612683-1613450
#start_codon TTG
CLASSIFICATION #superfamily hypothetical protein MJ1632
SUMMARY        #length 255 #molecular-weight 29795 #checksum 9201
               Query Match      58.7%; Score 61; DB 2; Length 255;
               Best Local Similarity 46.2%; Pred. No. 4.00e+00;
               Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
Db 193 KFLSNWDLPYRIT 205
   |||:|:|:|
QY 1 KFLAVWKITYKDT 13
```

Search completed: Wed Aug 4 15:30:42 1999
Job time : 8 secs.

This Page Blank (uspto)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Aug 4 15:30:59 1999; MasPar time 3.19 Seconds
Tabular output not generated. 115.309 Million cell updates/sec

Title: >US-09-049-847-4
Description: (1-13) from US09049847.pep
Perfect Score: 104
Sequence: 1 KFLAVWKIYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 27.515; Variance 38.609; scale 0.713

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	87	83.7	2205	1	POLG_POL1M	GENOME POLYPROTEIN [CO 7.73e-06
2	87	83.7	2208	1	POLH_POL1M	GENOME POLYPROTEIN [CO 7.73e-06
3	85	81.7	2205	1	POLG_POL2W	GENOME POLYPROTEIN [CO 2.11e-05
4	85	81.7	2207	1	POLG_POL2L	GENOME POLYPROTEIN [CO 2.11e-05
5	85	81.7	2209	1	POLG_POL1S	GENOME POLYPROTEIN [CO 2.11e-05
6	79	76.0	2206	1	POLG_POL3L	GENOME POLYPROTEIN [CO 4.01e-04
7	77	74.0	2206	1	POLG_POL32	GENOME POLYPROTEIN [CO 1.05e-03
8	70	67.3	1040	1	EG15_CAEEL	MYOBLAST GROWTH FACTOR 2.74e-02
9	64	61.5	2206	1	POLG_CXA21	GENOME POLYPROTEIN [CO 3.91e-01
10	62	59.6	874	1	DP3A_MYCGE	DNA POLYMERASE III, AL 9.20e-01
11	61	58.7	149	1	GLB3_MORMR	GLOBIN III. 1.40e+00
12	59	56.7	1938	1	MYSS_CHICK	MYOSIN HEAVY CHAIN, SK 3.21e+00
13	58	55.8	230	1	YX09_CAEEL	HYPOTHETICAL 26.5 KD P 4.81e+00
14	58	55.8	847	1	SYA_HELPY	ALANYL-TRNA SYNTHETASE 4.81e+00
15	58	55.8	874	1	POL1_HUMAN	RETROVIRUS-RELATED POL 4.81e+00
16	58	55.8	878	1	SYA_BACSU	ALANYL-TRNA SYNTHETASE 4.81e+00
17	58	55.8	1332	1	XKDO_BACSU	PHAGE-LIKE ELEMENT PBS 7.19e+00
18	57	54.8	241	1	YNBO_YEAST	HYPOTHETICAL 27.5 KD P 7.19e+00
19	57	54.8	396	1	RMAR_YEAST	MITOCHONDRIAL RIBOSOMA 7.19e+00
20	57	54.8	415	1	EF1G_YEAST	ELONGATION FACTOR 1-GA 7.19e+00
21	57	54.8	467	1	KCOT_MOUSE	COT PROTO-ONCOGENE SER 7.19e+00
22	57	54.8	644	1	VP4_BTIV1	VP4 CORE PROTEIN. 7.19e+00
23	57	54.8	644	1	VP4_BTIV3	VP4 CORE PROTEIN. 7.19e+00

24	57	54.8	654	1	VP4_BTIV10	VP4 CORE PROTEIN. 7.19e+00
25	57	54.8	1940	1	MYSE_CHICK	MYOSIN HEAVY CHAIN, FA 7.19e+00
26	56	53.8	151	1	CAV3_HUMAN	CAVEOLIN-3 (M-CAVEOLIN 1.07e+01
27	56	53.8	190	1	YCG8_YEAST	VERY HYPOTHETICAL 22.1 1.07e+01
28	56	53.8	346	1	YRAQ_ECOLI	HYPOTHETICAL 37.3 KD P 1.07e+01
29	56	53.8	467	1	KCOT_HUMAN	COT PROTO-ONCOGENE SER 1.07e+01
30	56	53.8	538	1	BUD5_YEAST	BUD SITE SELECTION PRO 1.07e+01
31	56	53.8	904	1	SYA_MYCIU	ALANYL-TRNA SYNTHETASE 1.07e+01
32	56	53.8	1934	1	MYSB_MESAU	MYOSIN HEAVY CHAIN, CA 1.07e+01
33	56	53.8	1935	1	MYSB_HUMAN	MYOSIN HEAVY CHAIN, CA 1.07e+01
34	56	53.8	1935	1	MYSB_PIG	MYOSIN HEAVY CHAIN, CA 1.07e+01
35	56	53.8	1938	1	MYSA_MOUSE	MYOSIN HEAVY CHAIN, CA 1.07e+01
36	56	53.8	1938	1	MYSA_RAT	MYOSIN HEAVY CHAIN, CA 1.07e+01
37	56	53.8	1939	1	MYSA_MESAU	MYOSIN HEAVY CHAIN, CA 1.07e+01
38	56	53.8	1939	1	MYSA_HUMAN	MYOSIN HEAVY CHAIN, CA 1.07e+01
39	56	53.8	2329	1	YS89_CAEEL	HYPOTHETICAL 254.3 KD 1.07e+01
40	56	53.8	4196	1	DYHC_SCHPO	DYNEIN HEAVY CHAIN, CY 1.07e+01
41	55	52.9	46	1	YPC4_ECOLI	HYPOTHETICAL 5.3 KD PR 1.58e+01
42	55	52.9	240	1	RESB_BACSU	TRANSCRIPTIONAL REGULA 1.58e+01
43	55	52.9	715	1	PERE_HUMAN	EOSINOPHIL PEROXIDASE 1.58e+01
44	55	52.9	860	1	SYL_ECOLI	LEUCYL-TRNA SYNTHETASE 1.58e+01
45	55	52.9	1096	1	PULA_KLEAE	PULLULANASE PRECURSOR 1.58e+01

ALIGNMENTS

RESULT 1
ID POLG_POL1M STANDARD; PRT; 2206 AA.
AC P03299;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)] (VERSION 1).
OS POLIOVIRUS TYPE 1 (STRAIN MAHONEY).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANCAK R., LEE J.J., VAN DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA.";
RL NATURE 291:547-553(1981).
RN [2]
RP SEQUENCE OF 1538-1573 FROM N.A.
RX MEDLINE; 81001866.
RA KITAMURA N., ADLER C.J., ROTHBERG P.G., MARTINKO J., NATHENSON S.G.,
RA WIMMER E.;
RT "The genome-linked protein of picornaviruses. VII. Genetic mapping of
RT poliovirus Vpg by protein and RNA sequence studies.";
RL CELL 21:295-302(1980).
RN [3]
RP SEQUENCE OF 1-68.
RX MEDLINE; 82242310.
RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
RT "Identification of the initiation site of poliovirus polypeptide
RT synthesis.";
RL J. VIROL. 42:1017-1028(1982).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE; 85300512.
RA HOGLE J.M., CHOW M., FILMAN D.J.;
RT "Three-dimensional structure of poliovirus at 2.9-A resolution.";
RL SCIENCE 229:1358-1365(1985).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE; 95120467.
RA GRANT R.A., HIREMATH C.N., FILMAN D.J., SYED R., ANDRIES K.,

RA HOGLE J.M.:
 RI "Structures of poliovirus complexes with anti-viral drugs:
 RT implications for viral stability and drug design.";
 RL CURR. BIOL. 4:784-797(1994).
 CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
 CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
 CC VP3, AND VP4.
 CC -!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -!- PTM: VPG IS LINKED, PROBABLY BY TYR-1545, TO THE URIDYLATE RESIDUE
 CC AT THE 5' END OF THE GENOME RNA. IT MAY PLAY A ROLE IN THE
 CC INITIATION OF RNA SYNTHESIS AND IT MAY ALSO BE INVOLVED IN
 CC MORPHOGENESIS.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
 CC -!- CAUTION: SEE ALSO VERSION 2 OF THIS PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V01148; G61237; -.
 DR PIR; A03897; GNNY1P.
 DR PDB; 2PLV; 15-JUL-93.
 DR PDB; 1FPT; 31-MAR-95.
 DR PDB; 1POV; 07-DEC-95.
 DR PDB; 1VBD; 11-JUL-96.
 DR PDB; 1AL2; 19-NOV-97.
 DR PDB; 1AR6; 03-DEC-97.
 DR PDB; 1AR7; 03-DEC-97.
 DR PDB; 1AR8; 03-DEC-97.
 DR PDB; 1AR9; 03-DEC-97.
 DR PDB; 1ASJ; 03-DEC-97.
 DR PDB; 1PO1; 03-DEC-97.
 DR PDB; 1PO2; 03-DEC-97.
 DR PFAM; PF00073; rhv; 3.
 DR PFAM; PF00548; Cys-protease-3C; 1.
 DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM; PF00910; RNA_helicase; 1.
 DR PFAM; PF00947; Pico_P2A; 1.
 KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
 KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
 KW 3D-STRUCTURE.
 FT INIT_MET 0 0
 FT CHAIN 1 68 COAT PROTEIN VP4.
 FT CHAIN 69 339 COAT PROTEIN VP2.
 FT CHAIN 340 577 COAT PROTEIN VP3.
 FT CHAIN 578 879 COAT PROTEIN VP1.
 FT CHAIN 880 1028 CORE PROTEIN P2A.
 FT CHAIN 1029 1125 CORE PROTEIN P2B.
 FT CHAIN 1126 1454 CORE PROTEIN P2C.
 FT CHAIN 1455 1541 CORE PROTEIN P3A.
 FT CHAIN 1542 1563 GENOME-LINKED PROTEIN VPG.
 FT CHAIN 1564 1745 PICORNAIN 3C.
 FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.
 FT LIPID 1 1 MYRISTATE.
 FT BINDING 1544 1544 URIDYLATE AT THE 5' END OF THE GENOME RNA.
 FT ACT_SITE 1710 1710 PROTEASE (POTENTIAL).
 FT STRAND 3 6
 FT STRAND 25 28
 FT HELIX 35 37
 FT STRAND 45 45
 FT TURN 49 49
 FT HELIX 50 53
 FT STRAND 56 56
 FT TURN 62 63
 FT TURN 74 75
 FT TURN 79 80
 FT STRAND 82 86

FT TURN 87 88
 FT STRAND 89 93
 FT STRAND 101 101
 FT HELIX 102 104
 FT TURN 112 114
 FT STRAND 122 122
 FT HELIX 125 127
 FT TURN 128 128
 FT STRAND 132 133
 FT STRAND 137 140
 FT TURN 141 142
 FT STRAND 146 150
 FT TURN 151 151
 FT HELIX 152 154
 FT TURN 155 156
 FT HELIX 158 166
 FT STRAND 167 179
 FT TURN 184 185
 FT STRAND 186 196
 FT TURN 197 197
 FT STRAND 202 202
 FT HELIX 213 216
 FT TURN 219 221
 FT STRAND 223 224
 FT STRAND 226 226
 FT TURN 235 236
 FT STRAND 242 242
 FT STRAND 245 245
 FT HELIX 246 248
 FT TURN 249 252
 FT HELIX 255 259
 FT STRAND 262 266
 FT TURN 267 269
 FT STRAND 272 277
 FT STRAND 286 286
 FT TURN 288 290
 FT STRAND 291 291
 FT STRAND 294 306
 FT STRAND 313 330
 FT STRAND 336 336
 FT TURN 347 350
 FT TURN 354 355
 FT STRAND 362 362
 FT TURN 365 366
 FT STRAND 378 379
 FT STRAND 381 381
 FT HELIX 382 386
 FT TURN 387 387
 FT STRAND 390 391
 FT STRAND 396 396
 FT TURN 398 402
 FT HELIX 404 407
 FT STRAND 409 412
 FT STRAND 422 425
 FT TURN 428 430
 FT TURN 432 436
 FT HELIX 438 443
 FT TURN 444 445
 FT STRAND 446 450
 FT STRAND 453 459
 FT TURN 463 464
 FT STRAND 466 466
 FT STRAND 468 474
 FT HELIX 484 488
 FT TURN 489 489
 FT STRAND 491 496
 FT STRAND 502 507
 FT STRAND 516 517
 FT HELIX 522 524
 FT STRAND 528 533
 FT STRAND 538 538
 FT TURN 541 542
 FT STRAND 546 555

FT TURN 557 558
FT STRAND 560 564
FT STRAND 599 599
FT TURN 600 601
FT STRAND 603 603
FT STRAND 607 607
FT TURN 619 620
FT STRAND 621 622

....
Note: remainder of annotations omitted.

Query Match 83.7%; Score 87; DB 1; Length 2206;
Best Local Similarity 84.6%; Pred. No. 7.73e-06;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 680 KLFAYWKITYKDT 692
|:|||||
QY 1 KFLAVWKITYKDT 13

RESULT 2
ID POLH_POL1M STANDARD; PRT; 2208 AA.
AC P03300; Q84879; Q84880; Q89679;
DT 21-JUL-1986 (REL. 01, CREATED)
DI 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DI 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C; P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.49)] (VERSION 2).
OS POLIOVIRUS TYPE 1 (STRAIN MAHONEY).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 82060159.
RA RACANIELLO V.R., BALTIMORE D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
RT complete nucleotide sequence of the viral genome."
RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE; 85300512.
RA HOGLE J.M., CHOW M., FILMAN D.J.;
RT "Three-dimensional structure of poliovirus at 2.9-A resolution."
RL SCIENCE 229:1358-1365(1985).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-880.
RX MEDLINE; 95120457.
RA GRANT R.A., HIREMAH C.N., FILMAN D.J., SYED R., ANDRIES K.,
RA HOGLE J.M.;
RT "Structures of poliovirus complexes with anti-viral drugs:
RT implications for viral stability and drug design."
RL CURR. BIOL. 4:784-797(1994).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHERAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMTIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -!- CAUTION: SEE ALSO VERSION 1 OF THIS PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; J02281; G332887; -.
DR EMBL; V01149; G61253; -.

DR EMBL; V01149; E9211; ALT_SEQ.
DR EMBL; V01149; E9212; ALT_SEQ.
DR EMBL; V01149; E9213; ALT_SEQ.
DR PIR; A03898; GNNY2P.
DR PDB; 2PLV; 15-JUL-93.
DR PDB; 1FPT; 31-MAR-95.
DR PDB; 1POV; 07-DEC-95.
DR PDB; 1VBD; 11-JUL-96.
DR PDB; 1AL2; 19-NOV-97.
DR PDB; 1AR6; 03-DEC-97.
DR PDB; 1AR7; 03-DEC-97.
DR PDB; 1AR8; 03-DEC-97.
DR PDB; 1AR9; 03-DEC-97.
DR PDB; 1ASJ; 03-DEC-97.
DR PDB; 1PO1; 03-DEC-97.
DR PDB; 1PO2; 03-DEC-97.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
KW 3D-STRUCTURE.
FT INIT_MEI 0 0
FT CHAIN 1 68 COAT PROTEIN VP4.
FT CHAIN 69 340 COAT PROTEIN VP2.
FT CHAIN 341 578 COAT PROTEIN VP3.
FT CHAIN 579 880 COAT PROTEIN VP1.
FT CHAIN 881 1029 CORE PROTEIN P2A.
FT CHAIN 1030 1126 CORE PROTEIN P2B.
FT CHAIN 1127 1455 CORE PROTEIN P2C.
FT CHAIN 1456 1542 CORE PROTEIN P3A.
FT CHAIN 1543 1564 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1565 1746 PICORNAIN 3C.
FT CHAIN 1747 2208 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 1 1 MYRISTATE.
FT ACT_SITE 1711 1711 PROTEASE (POTENTIAL).
FT ACT_SITE 1725 1725 PROTEASE (POTENTIAL).
FT STRAND 3 6
FT STRAND 25 28
FT HELIX 35 37
FT STRAND 45 45
FT TURN 49 49
FT HELIX 50 53
FT STRAND 56 56
FT TURN 62 63
FT TURN 74 75
FT TURN 79 80
FT STRAND 82 86
FT TURN 87 88
FT STRAND 89 93
FT STRAND 96 101
FT HELIX 102 104
FT TURN 112 114
FT STRAND 122 122
FT HELIX 125 127
FT TURN 128 128
FT STRAND 132 133
FT STRAND 137 140
FT TURN 141 142
FT STRAND 146 150
FT TURN 151 151
FT HELIX 152 154
FT TURN 155 156
FT HELIX 158 166
FT STRAND 167 179
FT TURN 184 185
FT STRAND 186 196
FT TURN 197 197
FT STRAND 202 202
FT HELIX 213 216
FT HELIX 219 221

223 STRAND 224
226 STRAND 226
235 TURN 236
242 STRAND 242
245 STRAND 245
246 HELIX 248
249 TURN 252
255 HELIX 260
263 STRAND 267
268 TURN 270
273 STRAND 278
287 STRAND 287
289 TURN 291
292 STRAND 292
295 STRAND 307
314 STRAND 331
337 STRAND 337
348 TURN 351
355 TURN 356
363 STRAND 363
366 TURN 367
379 STRAND 380
382 STRAND 382
383 HELIX 387
388 TURN 388
391 STRAND 391
397 STRAND 397
399 TURN 403
405 HELIX 408
410 STRAND 410
413 STRAND 426
423 STRAND 423
429 TURN 431
433 TURN 437
439 HELIX 443
445 TURN 444
447 STRAND 451
454 STRAND 460
464 TURN 465
467 STRAND 467
469 STRAND 475
485 HELIX 489
490 TURN 490
492 STRAND 497
503 STRAND 508
517 STRAND 518
523 HELIX 525
529 STRAND 534
539 STRAND 539
542 TURN 543
547 STRAND 547
558 TURN 558
561 STRAND 561
600 STRAND 600
601 TURN 602
604 STRAND 604
608 STRAND 608
620 TURN 621
622 STRAND 622
625 HELIX 627
635 STRAND 637
638 TURN 638
644 STRAND 644
649 STRAND 649
651 HELIX 653
654 STRAND 654
655 HELIX 655
663 STRAND 663
675 TURN 675
684 STRAND 684
695 STRAND 695
701 TURN 701
703 STRAND 703
732 STRAND 732

FT TURN 740 741
FT TURN 748 749
...
Note: remainder of annotations omitted.
Query Match 83.7%; Score 87; DB 1; Length 2208;
Best Local Similarity 84.6%; Pred. No. 7.73e-06;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 681 KLFAYWKITYKDT 693
QY 1 KFLAVWKITYKDT 13
RESULT 3
ID POLG_POL2W STANDARD; PRT; 2205 AA.
AC P23069;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS POLIOVIRUS TYPE 2 (STRAIN W-2).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90155230.
RA PEVEAR D.C., OH C.K., CUNNINGHAM L.L., CALENOFF M., JUBELT B.;
RT "Localization of genomic regions specific for the attenuated, mouse-
adapted poliovirus type 2 strain W-2.";
RL J. GEN. VIROL. 71:43-52(1990).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D00625; D1000971; ALT_SEQ.
PIR; A34032; GNNY2W.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
DR HSP; P03299; lPOV.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69 COAT PROTEIN VP4 (P1A).
FT CHAIN 70 340 COAT PROTEIN VP2 (P1B).
FT CHAIN 341 578 COAT PROTEIN VP3 (P1C).
FT CHAIN 579 879 COAT PROTEIN VP1 (P1D).
FT CHAIN 880 1028 PROTEASE 2A.
FT CHAIN 1029 1125 CORE PROTEIN 2B.
FT CHAIN 1126 1454 CORE PROTEIN 2C.
FT CHAIN 1455 1541 CORE PROTEIN 3A.
FT CHAIN 1542 1563 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1564 1746 PICORNAIN 3C.
FT CHAIN 1747 2205 RNA-DIRECTED RNA POLYMERASE 3D.
FT LIPID 2 2 MYRISTATE (BY SIMILARITY).
FT ACT_SITE 1710 1710 PROTEASE (POTENTIAL).

FT ACT_SITE 1724 1724 PROTEASE (POTENTIAL).
SQ SEQUENCE 2205 AA; 245701 MW; DA976BE8 CRC32;

Query Match 81.7%; Score 85; DB 1; Length 2205;
Best Local Similarity 76.9%; Pred. No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLF5VWKITYKDT 693
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 4
ID POLG_POL2L STANDARD; PRT; 2207 AA.
AC P06210;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
(EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
(EC 2.7.7.48)].
OS POLIOVIRUS TYPE 2 (STRAIN LANSING).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86115399.
RA LA MONICA N., MERIAM C., RACANIELLO V.R.;
RT "Mapping of sequences required for mouse neurovirulence of poliovirus
type 2 Lansing.";
RL J. VIROL. 57:515-525(1985).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M12197; G332891; -.
DR PIR; A29507; GNNY5P.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
DR HSSP; P03299; IPOV.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69
FT CHAIN 70 340
FT CHAIN 341 578
FT CHAIN 579 879
FT CHAIN 880 1028
FT CHAIN 1029 1125
FT CHAIN 1126 1454
FT CHAIN 1455 1541
FT CHAIN 1542 1563
FT CHAIN 1564 1746
FT CHAIN 1747 2207
FT LIPID 2
FT ACT_SITE 1710 1710
FT ACT_SITE 1724 1724
FT ACT_SEQUENCE 2207 AA; 245829 MW; 36A83506 CRC32;

Query Match 81.7%; Score 85; DB 1; Length 2207;
Best Local Similarity 76.9%; Pred. No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 681 KLF5VWKITYKDT 693
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 5
ID POLG_POLLS STANDARD; PRT; 2209 AA.
AC P03301;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
(EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
(EC 2.7.7.48)].
OS POLIOVIRUS TYPE 1 (STRAIN SABIN).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83299876.
RA NOMOTO A., OMATA T., TOYODA H., KUGE S., HORIE H., KATAOKA Y.,
RA GENBA Y., NAKANO Y., IMURA N.;
RT "Complete nucleotide sequence of the attenuated poliovirus Sabin 1
strain genome.";
RL PROC. NATL. ACAD. SCI. U.S.A. 79:5793-5797(1982).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- THIS VIRUS IS A LIVE VACCINE STRAIN DERIVED FROM THE MAHONEY
STRAIN BY SPONTANEOUS MUTATIONS DURING THE ATTENUATION PROCESS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; V01150; G61258; -.
DR PIR; A03899; GNNY3P.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
DR HSSP; P03299; IPOV.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69
FT CHAIN 70 341
FT CHAIN 342 579
FT CHAIN 580 881
FT CHAIN 882 1030
FT CHAIN 1031 1127
FT CHAIN 1128 1456
FT CHAIN 1457 1543
FT CHAIN 1544 1565
FT CHAIN 1566 1747
FT CHAIN 1748 2209
FT LIPID 2
FT ACT_SITE 1712 1712
FT ACT_SITE 1726 1726

SQ SEQUENCE 2209 AA; 246576 MW; 25407F3A CRC32;

Query Match 81.7%; Score 85; DB 1; Length 2209;
Best Local Similarity 76.9%; Pred. No. 2.11e-05;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 682 KLFTVWKITYKDT 694
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 6

ID POLG_POL3L STANDARD; PRT; 2206 AA.
AC P03302;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
(EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
(EC 2.7.7.48)].
OS POLIOVIRUS TYPE 3 (STRAINS P3/LEON/37 AND P3/LEON 12A[1]B).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3/LEON/37;
RX MEDLINE; 84170338.
RA STANWAY G., HUGHES P.J., MOUNTFORD R.C., REEVE P., MINOR P.D.,
RA SCHILD G.C., ALMOND J.W.;
RT "Comparison of the complete nucleotide sequences of the genomes of
the neurovirulent poliovirus P3/Leon/37 and its attenuated Sabin
vaccine derivative P3/Leon 12alb.";
RL PROC. NATL. ACAD. SCI. U.S.A. 81:1539-1543(1984).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P3/LEON 12A[1]B;
RX MEDLINE; 83299239.
RA STANWAY G., CANN A.J., HAUPTMANN R., HUGHES P., CLARKE L.D.,
RA MOUNTFORD R.C., MINOR P.D., SCHILD G.C., ALMOND J.W.;
RT "The nucleotide sequence of poliovirus type 3 leon 12 alb: comparison
with poliovirus type 1";
RL NUCLEIC ACIDS RES. 11:5629-5643(1983).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 1-878.
RX MEDLINE; 95120467.
RA GRANT R.A., HIREMATH C.N., FILMAN D.J., SYED R., ANDRIES K.,
RA HOGLE J.M.;
RT "Structures of poliovirus complexes with anti-viral drugs:
implications for viral stability and drug design.";
RL CURR. BIOL. 4:784-797(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1-878.
RA SYED R., FILMAN D.J., HOGLE J.M.;
RL SUBMITTED (MAR-1995) TO THE PDB DATA BANK.
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- THE SEQUENCE OF STRAIN SABIN VACCINE P3/LEON/37 IS SHOWN.
CC -!- THE STRAIN SABIN VACCINE P3/LEON/37 IS THE PROGENITOR OF THE
STRAIN SABIN VACCINE P3/LEON 12A[1]B.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; K01392; G332896; -.
DR EMBL; X00925; G61155; -.
DR PIR; A03900; GNNY4P.
DR PDB; 1PIV; 20-JUL-95.
DR PDB; 1PVC; 15-SEP-95.
DR PDB; 1VBA; 11-JUL-96.
DR PDB; 1VBB; 11-JUL-96.
DR PDB; 1VBC; 11-JUL-96.
DR PDB; 1VBE; 11-JUL-96.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00580; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
3D-STRUCTURE.
FT CHAIN 2 69 COAT PROTEIN VP4.
FT CHAIN 70 340 COAT PROTEIN VP2.
FT CHAIN 341 578 COAT PROTEIN VP3.
FT CHAIN 579 878 COAT PROTEIN VP1.
FT CHAIN 879 1027 CORE PROTEIN P2A.
FT CHAIN 1028 1124 CORE PROTEIN P2B.
FT CHAIN 1125 1453 CORE PROTEIN P2C.
FT CHAIN 1454 1540 CORE PROTEIN P3A.
FT CHAIN 1541 1562 GENOME-LINKED PROTEIN VPG.
FT CHAIN 1563 1745 PICORNAIN 3C.
FT CHAIN 1746 2206 RNA-DIRECTED RNA POLYMERASE P3D.
FT LIPID 2 2 MYRISTATE.
FT ACT_SITE 1709 1709 PROTEASE (POTENTIAL).
FT ACT_SITE 1723 1723 PROTEASE (POTENTIAL).
FT VARIANT 431 431 S -> F (IN P3/LEON 12A[1]B).
FT VARIANT 864 864 K -> R (IN P3/LEON 12A[1]B).
FT VARIANT 908 908 T -> A (IN P3/LEON 12A[1]B).
SQ SEQUENCE 2206 AA; 246163 MW; 78B79E4F CRC32;

Query Match 76.0%; Score 79; DB 1; Length 2206;
Best Local Similarity 69.2%; Pred. No. 4.01e-04;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 KLFAMWRITYKDT 691
|:::|||||

QY 1 KFLAVWKITYKDT 13

RESULT 7

ID POLG_POL32 STANDARD; PRT; 2206 AA.
AC P06209;
DT 01-JAN-1988 (REL. 06, CREATED)
DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
(EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
(EC 2.7.7.48)].
OS POLIOVIRUS TYPE 3 (STRAIN 23127).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87010550.
RA HUGHES P.J., EVANS D.M.A., MINOR P.D., SCHILD G.C., ALMOND J.W.,
RA STANWAY G.;
RT "The nucleotide sequence of a type 3 poliovirus isolated during a
recent outbreak of poliomyelitis in Finland.";
RL J. GEN. VIROL. 67:2093-2102(1986).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X04468; G61113; -.
DR PIR; A27245; GNNY27.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
DR HSP; P03302; lPiv.
KW POLYPROTEIN; COAT PROTEIN; CORE PROTEIN; TRANSFERASE;
KW RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION.
FT CHAIN 2 69
FT CHAIN 70 340
FT CHAIN 341 578
FT CHAIN 579 878
FT CHAIN 879 1027
FT CHAIN 1028 1124
FT CHAIN 1125 1453
FT CHAIN 1454 1540
FT CHAIN 1541 1562
FT CHAIN 1563 1745
FT CHAIN 1745 2206
FT LIPID 2 2
FT ACT_SITE 1709 1709
FT ACT_SITE 1723 1723
SQ SEQUENCE 2206 AA; 245731 MW; 245809F5 CRC32;

Query Match 74.0%; Score 77; DB 1; Length 2206;
Best Local Similarity 59.2%; Pred. No. 1.05e-03;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 679 KLFATWRTYKDT 691
QY 1 KFLAVWKITYKDT 13
RESULT 8
ID EG15_CAEEL STANDARD; PRT; 1040 AA.
AC Q10656;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MYOBLAST GROWTH FACTOR RECEPTOR ELG-15 PRECURSOR (EC 2.7.1.112).
GN EGL-15.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE; 96069862.
RA DEVORE D.L., HORVITZ H.R., STERN M.J.;
RT "An FGF receptor signaling pathway is required for the normal cell
RT migrations of the sex myoblasts in C. elegans hermaphrodites.";
RL CELL 83:611-620(1995).
CC -!- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED FOR
CC THE GUIDANCE OF SEX MYOBLAST MIGRATION DURING GONAD DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC EXTRACELLULAR DOMAIN CONTAINS 3 IG-LIKE DOMAINS.
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U39761; G1079712; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00047; ig; 3.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P11362; lFGI.
KW RECEPTOR; GLYCOPROTEIN; TYROSINE-PROTEIN KINASE; ATP-BINDING;
KW TRANSFERASE; PHOSPHORYLATION; TRANSMEMBRANE; SIGNAL;
KW IMMUNOGLOBULIN FOLD.
FT SIGNAL 1 19
FT CHAIN 20 1040
FT DOMAIN 20 525
FT TRASMEM 526 549
FT DOMAIN 550 1040
FT DOMAIN 48 116
FT DOMAIN 307 374
FT DOMAIN 407 492
FT DOMAIN 640 931
FT NP_BIND 646 654
FT BINDING 672 672
FT ACT_SITE 797 797
FT MOD_RES 828 828
FT DISULFID 55 109
FT DISULFID 314 367
FT DISULFID 414 485
FT CARBOHYD 121 121
FT CARBOHYD 280 280
FT CARBOHYD 299 299
FT CARBOHYD 401 401
FT CARBOHYD 407 407
FT CARBOHYD 433 433
FT CARBOHYD 440 440
FT CARBOHYD 449 449
FT CARBOHYD 474 474
FT CARBOHYD 497 497
FT CARBOHYD 680 680
FT MUTAGEN 714 714
FT MUTAGEN 930 930
SQ SEQUENCE 1040 AA; 118956 MW; 97C19793 CRC32;

Query Match 67.3%; Score 70; DB 1; Length 1040;
Best Local Similarity 66.7%; Pred. No. 2.74e-02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 651 FGEVWKATYKET 662
QY 2 FLAVWKITYKDT 13
RESULT 9
ID POLG_CXA21 STANDARD; PRT; 2206 AA.
AC P22055;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN [CONTAINS: COAT PROTEINS VP1 TO VP4; CORE PROTEINS
DE P2A TO P2C, P3A; GENOME-LINKED PROTEIN VPG; PICORNAIN 3C
DE (EC 3.4.22.28) (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE P3D
DE (EC 2.7.7.48)].
OS COXSACKIEVIRUS A21 (STRAIN COE).
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90063544.

DT 01-JAN-1990 (REL. 13, CREATED)

DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)

DI 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)

DE MYOSIN HEAVY CHAIN, SKELETAL MUSCLE, ADULT.

OS GALLUS GALLUS (CHICKEN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;

OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.

RN [1]

RP SEQUENCE OF 1-205.

RC TISSUE=PECTORALIS MUSCLE;

RX MEDLINE; 92041767.

RA HAYASHIDA M., MAITA T., MATSUDA G.;

RT "The primary structure of skeletal muscle myosin heavy chain: I.

RT Sequence of the amino-terminal 23 kDa fragment.";

RL J. BIOCHEM. 110:54-59(1991).

RN [2]

RP SEQUENCE OF 206-636.

RC TISSUE=PECTORALIS MUSCLE;

RX MEDLINE; 92041768.

RA KOMINE Y., MAITA T., MATSUDA G.;

RT "The primary structure of skeletal muscle myosin heavy chain: II.

RT Sequence of the 50 kDa fragment of subfragment-1.";

RL J. BIOCHEM. 110:60-67(1991).

RN [3]

RP SEQUENCE OF 637-837.

RC TISSUE=PECTORALIS MUSCLE;

RX MEDLINE; 92041769.

RA MAITA T., MIYANISHI T., MATSUZONO K., TANIOKA Y., MATSUDA G.;

RT "The primary structure of skeletal muscle myosin heavy chain: III.

RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50

RT kDa, and 22 kDa fragments.";

RL J. BIOCHEM. 110:68-74(1991).

RN [4]

RP SEQUENCE OF 838-1938.

RC TISSUE=PECTORALIS MUSCLE;

RX MEDLINE; 92041770.

RA MAITA T., YAJIMA E., NAGATA S., MIYANISHI T., NAKAYAMA S., MATSUDA G.;

RT "The primary structure of skeletal muscle myosin heavy chain: IV.

RT Sequence of the rod, and the complete 1,938-residue sequence of the

RT heavy chain.";

RL J. BIOCHEM. 110:75-87(1991).

RN [5]

RP PRELIMINARY SEQUENCE OF 1-808.

RX MEDLINE; 87092420.

RA MAITA T., HAYASHIDA M., TANIOKA Y., KOMINE Y., MATSUDA G.;

RT "The primary structure of the myosin head.";

RL PROC. NATL. ACAD. SCI. U.S.A. 84:416-420(1987).

RN [6]

RP SEQUENCE OF 842-1270.

RX MEDLINE; 90121764.

RA WATANABE B.;

RT "Complete amino-acid sequence of subfragment-2 in adult chicken

RT skeletal muscle myosin.";

RL BIOL. CHEM. HOPPE-SEYLER 370:1027-1034(1989).

RN [7]

RP SEQUENCE OF 852-1108.

RX MEDLINE; 89374803.

RA WATANABE B.;

RT "Amino-acid sequence of the short subfragment-2 in adult chicken

RT skeletal muscle myosin.";

RL BIOL. CHEM. HOPPE-SEYLER 370:549-558(1989).

RN [8]

RP SEQUENCE OF 1145-1270.

RX MEDLINE; 89228549.

RA WATANABE B.;

RT "Amino-acid sequence of the hinge region in chicken myosin

RT subfragment-2.";

RL BIOL. CHEM. HOPPE-SEYLER 370:55-61(1989).

RN [9]

RP SEQUENCE OF 1857-1938 FROM N.A.

RX MEDLINE; 87217964.

RA MORIARTY D.M., BARRINGER K.J., DODGSON J.B., RICHTER H.E.,

RA YOUNG R.B.;

RT "Genomic clones encoding chicken myosin heavy-chain genes.";

RL DNA 6:91-99(1987).

RN [10]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 1-843.

RX MEDLINE; 93303624.

RA RAYMENT I., RYPIEWSKI W.R., SCHMIDT-BASE K., SMITH R.,

RA TOMCHICK D.R., BENNING M.M., WINKELMANN D.A., WESENBERG G.,

RA HOLDEN H.M.;

RT "Three-dimensional structure of myosin subfragment-1: a molecular

RT motor.";

RL SCIENCE 261:50-58(1993).

CC -!- FUNCTION: MUSCLE CONTRACTION. MYOSIN IS A PROTEIN THAT BINDS TO

CC F-ACTIN AND HAS ATPASE ACTIVITY THAT IS ACTIVATED BY F-ACTIN.

CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -!- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.

CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -!- EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMM)

CC AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE SPLIT FURTHER INTO

CC 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).

CC -!- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,

CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE

CC CONSERVED.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M16557; G212372; -.

PIR; PX0050; PX0051.

PIR; A26821; A26821.

PIR; S02082; S02082.

PIR; S04501; S04501.

PIR; S05515; S05515.

PDB; 2MYS; 11-JAN-97.

PFAM; PF00063; myosin_head; 1.

PFAM; PF00612; IQ; 1.

KW MYOSIN; MUSCLE PROTEIN; COILED COIL; THICK FILAMENT; ACTIN-BINDING;

KW ATP-BINDING; METHYLATION; ALKYLATION; PHOSPHORYLATION; ACETYLATION;

KW HEPTAD REPEAT PATTERN; MULTIPEPTIDE FAMILY; 3D-STRUCTURE.

FT DOMAIN 1 837

FT DOMAIN 838 840

FT DOMAIN 841 1880

FT DOMAIN 848 1289

FT DOMAIN 1290 1303

FT DOMAIN 1304 1880

FT DOMAIN 841 1938

FT NP_BIND 179 186

FT DOMAIN 657 679

FT DOMAIN 759 773

FT MOD_RES 1 1

FT MOD_RES 35 35

FT MOD_RES 130 130

FT MOD_RES 551 551

FT MOD_RES 755 755

FT MOD_RES 697 697

FT MOD_RES 707 707

FT CONFLICT 907 907

FT CONFLICT 1863 1863

FT CONFLICT 1929 1931

FT SEQUENCE 1938 AA; 222972 MW; 2ACE77FE CRC32;

Query Match 56.7%; Score 59; DB 1; Length 1938;

Best Local Similarity 53.8%; Pred. No. 3.21e+00;


```
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
Db 1453 KILAEWKQKYEET 1465
   ||| ||| | :|
QY 1 KFLAVWKITYKDT 13

RESULT 13
ID YX09.CABEL STANDARD; PRT; 230 AA.
AC Q11115;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 26.5 KD PROTEIN C03B1.9 IN CHROMOSOME X.
GN C03B1.9.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA MARTIN J.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U40952; G1072239; -.
DR WORMPEP; C03B1.9; CE03910.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 230 AA; 26473 MW; 0F82E358 CRC32;

Query Match 55.8%; Score 58; DB 1; Length 230;
Best Local Similarity 50.0%; Pred. No. 4.81e+00;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 74 KFLTKWRTVY 83
   ||| :| :|
QY 1 KFLAVWKITY 10

RESULT 14
ID SYA.HELPY STANDARD; PRT; 847 AA.
AC P56452;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE ALANYL-TRNA SYNTHETASE (EC 6.1.1.7) (ALANINE--TRNA LIGASE) (ALARS).
GN ALAS OR HP1241.
OS HELICOBACTER PYLORI (CAMPYLOBACTER PYLORI).
OC BACTERIA; PROTEOBACTERIA; EPSILON SUBDIVISION; HELICOBACTER GROUP;
OC HELICOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE; 97394467.
RA TOME J.-F., WHITE O., KERLAVAGE A.R., CLAYTON R.A., SUTTON G.G.,
RA FLEISCHMANN R.D., KETCHUM K.A., KLENK H.-P., GILL S., DOUGHERTY B.A.,
RA NELSON K., QUACKENBUSH J., ZHOU L., KIRKNESS E.F., PETERSON S.,
RA LOFTUS B., RICHARDSON D., DODSON R., KHALAK H.G., GLODEK A.,
RA MCKENNEY K., FITZGERALD L.M., LEE N., ADAMS M.D., HICKEY E.K.,
RA BERG D.E., GOCAYNE J.D., UTTERBACK T.R., PETERSON J.D., KELLEY J.M.,
RA COTTON M.D., WEIDMAN J.M., FUJII C., BOWMAN C., WATTHEY L., WALLIN E.,
RA HAYES W.S., BORODOVSKY M., KARP P.D., SMITH H.O., FRASER C.M.,
RA VENTER J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RI pylori.";
RL NATURE 388:539-547(1997).
```

```
CC -!- CATALYTIC ACTIVITY: ATP + L-ALANINE + TRNA(ALA) = AMP +
CC PYROPHOSPHATE + L-ALANYL-TRNA(ALA).
CC -!- COFACTOR: BINDS A ZINC ION (BY SIMILARITY).
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000629; G2314404; -.
DR TIGR; HP1241; -.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; 1.
KW AMINOACYL-TRNA SYNTHETASE; PROTEIN BIOSYNTHESIS; LIGASE; ATP-BINDING.
SQ SEQUENCE 847 AA; 94700 MW; 3C3AAD24 CRC32;

Query Match 55.8%; Score 58; DB 1; Length 847;
Best Local Similarity 25.0%; Pred. No. 4.81e+00;
Matches 3; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Db 191 RELEIWNLVFMQ 202
   ||| :| :| :|
QY 1 KFLAVWKITYKD 12

RESULT 15
ID POLL_HUMAN STANDARD; PRT; 874 AA.
AC P10266;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RETROVIRUS-RELATED POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE
DE (EC 2.7.7.49); ENDONUCLEASE].
GN POL.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87036922.
RA ONO M., YASUNAGA T., MIYATA T., USHIKUBO H.;
RT "Nucleotide sequence of human endogenous retrovirus genome related to
RT the mouse mammary tumor virus genome.";
RL J. VIROL. 60:589-598(1986).
DR PIR; D24483; GNHUEP.
DR PFAM; PF00075; rnaaseH; 1.
DR PFAM; PF00078; rvt; 1.
DR PFAM; PF00552; integrase; 1.
DR PFAM; PF00665; rve; 1.
DR HSSP; P03366; 1HMV.
KW HYDROLASE; TRANSFERASE; RNA-DIRECTED DNA POLYMERASE; NUCLEASE;
KW ENDONUCLEASE; POLYPROTEIN.
FT CHAIN 36 250 REVERSE TRANSCRIPTASE.
FT CHAIN 585 764 ENDONUCLEASE.
SQ SEQUENCE 874 AA; 98936 MW; ADFE3749 CRC32;

Query Match 55.8%; Score 58; DB 1; Length 874;
Best Local Similarity 66.7%; Pred. No. 4.81e+00;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 722 KFLSQWKIS 730
   ||| :| :| :|
QY 1 KFLAVWKIT 9

Search completed: Wed Aug 4 15:31:05 1999
Job time : 6 secs.
```



```
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA.";
RL NATURE 291:547-553(1981).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82060159.
RA RACANIELLO V.R., BALTIMORE D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
RT complete nucleotide sequence of the viral genome.";
RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82216986.
RA EMINI E.A., ELZINGA M., WIMMER E.;
RT "Carboxy-terminal analysis of poliovirus proteins: termination of
RT poliovirus RNA translation and location of unique poliovirus
RT polypeptide cleavage sites.";
RL J. VIROL. 42:194-199(1982).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82242310.
RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
RT "Identification of the initiation site of poliovirus polyprotein
RT synthesis.";
RL J. VIROL. 42:1017-1028(1982).
DR EMBL; V01148; E9201; -.
DR PFAM; PF00073; rhv; 1.
SQ SEQUENCE 302 AA; 33523 MW; C199C130 CRC32;

Query Match 83.7%; Score 87; DB 14; Length 302;
Best Local Similarity 84.6%; Pred. No. 2.19e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 103 KLEAVWKITYKDT 115
QY 1 KFLAVWKITYKDT 13

RESULT 3
ID Q84866 PRELIMINARY; PRT; 879 AA.
AC Q84866;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DI 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME OF HUMAN POLIOVIRUS TYPE 1 (MAHONEY STRAIN)
DE (ONE OF TWO VERSIONS.).
OS POLIOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA.";
RL NATURE 291:547-553(1981).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
```

```
RX MEDLINE; 82060159.
RA RACANIELLO V.R., BALTIMORE D.;
RT "Molecular cloning of poliovirus cDNA and determination of the
RT complete nucleotide sequence of the viral genome.";
RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82216986.
RA EMINI E.A., ELZINGA M., WIMMER E.;
RT "Carboxy-terminal analysis of poliovirus proteins: termination of
RT poliovirus RNA translation and location of unique poliovirus
RT polypeptide cleavage sites.";
RL J. VIROL. 42:194-199(1982).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 82242310.
RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
RT "Identification of the initiation site of poliovirus polyprotein
RT synthesis.";
RL J. VIROL. 42:1017-1028(1982).
DR EMBL; V01148; E9174; -.
DR PFAM; PF00073; rhv; 3.
SQ SEQUENCE 879 AA; 97251 MW; 9F584E23 CRC32;

Query Match 83.7%; Score 87; DB 14; Length 879;
Best Local Similarity 84.6%; Pred. No. 2.19e-05;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 680 KLEPAVKITYKDT 692
QY 1 KFLAVWKITYKDT 13

RESULT 4
ID Q84865 PRELIMINARY; PRT; 2221 AA.
AC Q84865;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DI 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE GENOME OF HUMAN POLIOVIRUS TYPE 1 (MAHONEY STRAIN)
DE (ONE OF TWO VERSIONS.).
OS POLIOVIRUS TYPE 1.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE OF 1893-2221 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81013850.
RA KITAMURA N., WIMMER E.;
RT "Sequence of 1050 3'-terminal nucleotides of poliovirus RNA as
RT determined by a modification of the dideoxynucleotide method.";
RL PROC. NATL. ACAD. SCI. U.S.A. 77:3196-3200(1980).
RN [2]
RP SEQUENCE OF 1553-1609 FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81001866.
RA KITAMURA N., ADLER C.J., ROTHBERG P.G., MARTINKO J., NATHENSON S.G.,
RA WIMMER E.;
RT "The genome-linked protein of picornaviruses. VII. Genetic mapping of
RT poliovirus VPg by protein and RNA sequence studies.";
RL CELL 21:295-302(1980).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAHONEY STRAIN;
RX MEDLINE; 81220953.
RA KITAMURA N., SEMLER B.L., ROTHBERG P.G., LARSEN G.R., ADLER C.J.,
RA DORNER A.J., EMINI E.A., HANECAK R., LEE J.J., DER WERF S.,
RA ANDERSON C.W., WIMMER E.;
RT "Primary structure, gene organization and polypeptide expression of
RT poliovirus RNA.";
RL NATURE 291:547-553(1981).
```

[4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=MAHONEY STRAIN;
 RC MEDLINE: 82060159.
 RX RACANELLO V.R., BALTIMORE D.;
 RA "Molecular cloning of poliovirus cDNA and determination of the
 RT complete nucleotide sequence of the viral genome.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 78:4887-4891(1981).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAHONEY STRAIN;
 RX MEDLINE: 82216986.
 RA EMINI E.A., ELZINGA M., WIMMER E.;
 RT "Carboxy-terminal analysis of poliovirus proteins: termination of
 RT poliovirus RNA translation and location of unique poliovirus
 RL J. VIROL. 42:194-199(1982).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAHONEY STRAIN;
 RX MEDLINE: 82242310.
 RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
 RT "Identification of the initiation site of poliovirus polypeptide
 RT synthesis.";
 RL J. VIROL. 42:1017-1028(1982).
 DR EMBL: V01148; E9172; -.
 DR PFAM: PF00073; rhv; 3.
 DR PFAM: PF00548; Cys-protease-3C; 1.
 DR PFAM: PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM: PF00910; RNA_helicase; 1.
 DR PFAM: PF00947; Pico_P2A; 1.
 SQ SEQUENCE 2221 AA; 248205 MW; A4BEA672 CRC32;

Query Match 83.7%; Score 87; DB 14; Length 2221;
 Best Local Similarity 84.6%; Pred. No. 2.19e-05;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 695 KLFVWKITYKDT 707
 QY 1 KFLAVWKITYKDT 13

RESULT 5
 ID 091261 PRELIMINARY; PRT; 100 AA.
 AC 091261;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE VP1 PROTEIN (FRAGMENT).
 GN VP1.
 OS HUMAN POLIOVIRUS 1.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 OC ENTEROVIRUS.

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=3786ALB96;
 RX MEDLINE: 98312955.
 RA FIORE L., GENOVESE D., DIAMANTI E., CATONE S., RIDOLFI B.,
 RA IBRAHIMI B., KONOMI R., VAN DER AVOORT H.G., HOVI T., CRAINIC R.,
 RA SIMEONI P., AMATO C.;
 RT "Antigenic and Molecular Characterization of Wild Type 1 Poliovirus
 RT Causing Outbreaks of Poliomyelitis in Albania and Neighboring
 RT Countries in 1996.";
 RL J. CLIN. MICROBIOL. 36:1912-1918(1998).
 DR EMBL: AJ007966; E1311747; -.
 FT NON_TER 1
 FT NON_TER 100 100
 SQ SEQUENCE 100 AA; 10695 MW; 736C541C CRC32;

Query Match 81.7%; Score 85; DB 14; Length 100;
 Best Local Similarity 76.9%; Pred. No. 5.81e-05;
 Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 86 KLESVWKITYKDI 98
 QY 1 KFLAVWKITYKDT 13

RESULT 6
 ID 084884 PRELIMINARY; PRT; 302 AA.
 AC 084884;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GENOME OF HUMAN POLIOVIRUS, STRAIN SABIN 1.
 OS POLIOVIRUS.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 OC ENTEROVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SABIN 1;
 RX MEDLINE: 83299876.
 RA NOMOTO A., OMATA T., TOYODA H., KUGE S., HORIE H., KATAOKA Y.,
 RA GENBA Y., NAKANO Y., IMURA N.;
 RT "Complete nucleotide sequence of the attenuated poliovirus Sabin 1
 RT strain genome.";
 RL PROC. NATL. ACAD. SCI. U.S.A. 79:5793-5797(1982).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SABIN 1;
 RX MEDLINE: 82216986.
 RA EMINI E.A., ELZINGA M., WIMMER E.;
 RT "Carboxy-terminal analysis of poliovirus proteins: termination of
 RT poliovirus RNA translation and location of unique poliovirus
 RL J. VIROL. 42:194-199(1982).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SABIN 1;
 RX MEDLINE: 82242310.
 RA DORNER A.J., DORNER L.F., LARSEN G.R., WIMMER E., ANDERSON C.W.;
 RT "Identification of the initiation site of poliovirus polypeptide
 RT synthesis.";
 RL J. VIROL. 42:1017-1028(1982).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SABIN 1;
 RA NOMOTO A.;
 RL SUBMITTED (APR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL: V01150; E9218; -.
 DR PFAM: PF00073; rhv; 1.
 SQ SEQUENCE 302 AA; 33454 MW; 20DC41FA CRC32;

Query Match 81.7%; Score 85; DB 14; Length 302;
 Best Local Similarity 76.9%; Pred. No. 5.81e-05;
 Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Db 103 KLFVWKITYKDT 115
 QY 1 KFLAVWKITYKDT 13

RESULT 7
 ID 089966 PRELIMINARY; PRT; 912 AA.
 AC 089966;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE POLYPROTEIN (FRAGMENT).
 OS HUMAN POLIOVIRUS 1.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
 OC ENTEROVIRUS.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=97-30515 (16/97/182);
 RA MULDER M.N., REIMERINK J.H.J., STENVIK M., VAN DER AVOORT H.G.A.M.,

RA HOVI T., KOOPMANS M.P.G.;
RT "A Sabin Vaccine-Derived Field Isolate of Poliovirus Type 1
RI Displaying Aberrant Phenotypic and Genotypic Features, Including a
RI Deletion in Antigenic Site 1.";
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF065158; G3511177; -.
KW POLYPROTEIN.
FT NON_TER 912 912
SQ SEQUENCE 912 AA; 100953 MW; 6C3044FD CRC32;

Query Match
Best Local Similarity 81.7%; Score 85; DB 14; Length 912;
Matches 10; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Db 680 KLFVWKITYKDT 692
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 8
ID Q98595 PRELIMINARY; PRT; 2207 AA.
AC Q98595;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PV2 POLYPROTEIN.
OS HUMAN POLIOVIRUS 2.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84216300.
RA TOYODA H., KOHARA M., KATAOKA Y., SUGANUMA T., OMATA T., IMURA N.,
RA NOMOTO A.;
RT "Complete nucleotide sequences of all three poliovirus serotype
RT genomes. Implication for genetic relationship, gene function and
RT antigenic determinants.";
RL J. MOL. BIOL. 174:561-585(1984).
RN [2]
RP SEQUENCE FROM N.A.
RA NOMOTO A.;
RL SUBMITTED (APR-1988) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X00595; E275415; -.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
KW POLYPROTEIN; MEMBRANE.
FT CHAIN 1 879 POTENTIAL.
FT CHAIN 1 69 POTENTIAL.
FT CHAIN 70 340 POTENTIAL.
FT CHAIN 341 578 POTENTIAL.
FT CHAIN 579 879 POTENTIAL.
FT CHAIN 880 1454 POTENTIAL.
FT CHAIN 1029 1454 POTENTIAL.
FT CHAIN 1126 1454 POTENTIAL.
FT CHAIN 1455 2207 POTENTIAL.
FT CHAIN 1541 1562 POTENTIAL.
FT CHAIN 1564 2207 POTENTIAL.
FT CHAIN 1747 2207 POTENTIAL.
SQ SEQUENCE 2207 AA; 245967 MW; 709F83D9 CRC32;

Query Match
Best Local Similarity 78.8%; Score 82; DB 14; Length 2207;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 681 RLFSVWKITYKDT 693
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 8

ID Q84791 PRELIMINARY; PRT; 300 AA.
AC Q84791;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE (LEON) FRAGMENT ENCODING VP1 (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 83141766.
RA MINOR P.D., SCHILD G.C., BOOTMAN J., EVANS D.M.A., FERGUSON M.,
RA REEVE P., SPITZ M., STANWAY G., CANN A.J., HAUPTMANN R., CLARKE L.D.,
RA MOUNTFORD R.C., ALMOND J.W.;
RT "Location and primary structure of a major antigenic site for
RT poliovirus neutralization.";
RL NATURE 301:674-679(1983).
DR EMBL; V01132; G929810; -.
DR PFAM; PF00073; rhv; 1.
FT NON_TER 1 1
FT NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;
Query Match
Best Local Similarity 76.0%; Score 79; DB 14; Length 300;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
|:::|||||
QY 1 KFLAVWKITYKDT 13

RESULT 10
ID Q84891 PRELIMINARY; PRT; 300 AA.
AC Q84891;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE VP1 (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84249506.
RA MINOR P.D., EVANS D.M.A., SCHILD G.C., FERGUSON M., ALMOND J.W.;
RT "Identification of an antigenic site in the neutralization of type 3
RT poliovirus.";
RL REV. INFECT. DIS. 6:516-518(1984).
RN [2]
RP SEQUENCE OF 12-300 FROM N.A.
RX MEDLINE; 84249500.
RA ALMOND J.W., CANN A.J., MINOR P.D., REEVE P., SCHILD G.C.,
RA HAUPTMANN R., STANWAY G.;
RT "Nucleotide sequence from neurovirulent and attenuated strains of
RT type 3 poliovirus.";
RL REV. INFECT. DIS. 6:487-493(1984).
DR EMBL; M37321; G332982; -.
DR PFAM; PF00073; rhv; 1.
FT NON_TER 1 1
FT NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;

Query Match
Best Local Similarity 76.0%; Score 79; DB 14; Length 300;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 101 KLFAMWRITYKDT 113
|:::|||||
QY 1 KFLAVWKITYKDT 13


```
RESULT 11
ID Q84793 PRELIMINARY; PRT; 300 AA.
AC Q84793;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POLIOVIRUS P3/LEON/37 GENOME FRAGMENT ENCODING THE VP1 PROTEIN
DE (FRAGMENT).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LEON;
RX MEDLINE; 84004370.
RA STANWAY G., CANN A.J., HAUPTMANN R., MOUNTFORD R.C., CLARKE L.D.,
RA REEVE P., MINOR P.D., SCHILD G.C., ALMOND J.W.;
RT "Nucleic acid sequence of the region of the genome encoding capsid
RT protein VP1 of neurovirulent and attenuated type 3 polioviruses.";
RL EUR. J. BIOCHEM. 135:529-533(1983).
DR EMBL; V01540; G929811; -.
DR PFAM; PF00073; rhv; 1.
DR NON_TER 1
DR NON_TER 300 300
SQ SEQUENCE 300 AA; 33458 MW; 71308C5C CRC32;

Query Match 76.0%; Score 79; DB 14; Length 300;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDT 13

RESULT 12
ID Q84784 PRELIMINARY; PRT; 1628 AA.
AC Q84784;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPLETE SEQUENCE (STRAIN P3/119).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P3/119;
RX MEDLINE; 85037944.
RA CANN A.J., STANWAY G., HUGHES P.J., MINOR P.D., EVANS D.M.A.,
RA SCHILD G.C., ALMOND J.W.;
RT "Reversion to neurovirulence of the live-attenuated Sabin type 3 oral
RT poliovirus vaccine.";
RL NUCLEIC ACIDS RES. 12:7787-7792(1984).
DR EMBL; X01076; E9008; -.
DR PFAM; PF00073; rhv; 1.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
SQ SEQUENCE 1628 AA; 182445 MW; 53EE6A08 CRC32;

Query Match 76.0%; Score 79; DB 14; Length 1628;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 101 KLFAMWRITYKDT 113
QY 1 KFLAVWKITYKDT 13

RESULT 13
ID Q84792 PRELIMINARY; PRT; 2206 AA.
AC Q84792;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PV3 POLYPROTEIN.
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 84216300.
RA TOYODA H., KOHARA M., KATAOKA Y., SUGANUMA T., OMATA T., IMURA N.,
RA NOMOTO A.;
RT "Complete nucleotide sequences of all three poliovirus serotype
RT genomes. Implication for genetic relationship, gene function and
RT antigenic determinants.";
RL J. MOL. BIOL. 174:561-585(1984).
RN [2]
RP SEQUENCE FROM N.A.
RA NOMOTO A.;
RL SUBMITTED (APR-1985) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; X00596; G61140; -.
DR PFAM; PF00073; rhv; 3.
DR PFAM; PF00548; Cys-protease-3C; 1.
DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR PFAM; PF00910; RNA_helicase; 1.
DR PFAM; PF00947; Pico_P2A; 1.
DR POLYPROTEIN.
KW POLYPROTEIN.
SQ SEQUENCE 2206 AA; 246295 MW; 64B5536A CRC32;

Query Match 76.0%; Score 79; DB 14; Length 2206;
Best Local Similarity 69.2%; Pred. No. 1.03e-03;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 KLFAMWRITYKDT 691
QY 1 KFLAVWKITYKDT 13

RESULT 14
ID Q84892 PRELIMINARY; PRT; 878 AA.
AC Q84892;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEINS (FRAGMENTS).
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAUKETT COP;
RX MEDLINE; 94160574.
RA HUOVILAINEN A., KINNUNEN L., POYRY T., LAAKSONEN L., ROIVAINEN M.,
RA HOVI T.;
RT "Poliovirus type 3/Saukett; antigenic and structural correlates of
RT sequence variation in the capsid proteins.";
RL VIROLOGY 199:228-232(1994).
DR EMBL; L23844; G388329; -.
DR PFAM; PF00073; rhv; 3.
DR CHAIN 1 >69 CAPSID PROTEIN.
FT NON_CONS 69 70
FT CHAIN 70 >340 CAPSID PROTEIN.
FT NON_CONS 340 341
FT CHAIN 341 >578 CAPSID PROTEIN.
FT NON_CONS 578 579 CAPSID PROTEIN.
FT CHAIN 579 >878 CAPSID PROTEIN.
FT NON_TER 878 878
SQ SEQUENCE 878 AA; 97326 MW; 1474D049 CRC32;

Query Match 74.0%; Score 77; DB 14; Length 878;
Best Local Similarity 69.2%; Pred. No. 2.63e-03;
```


Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

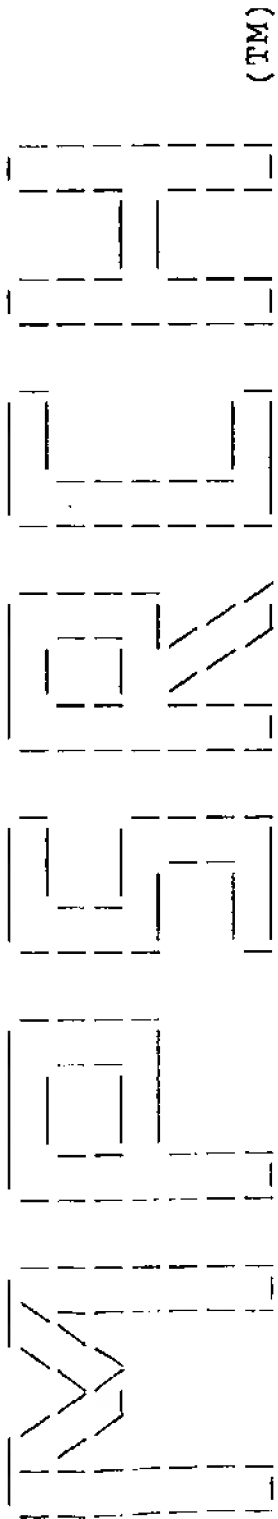
Db 679 KLEATWRITYKDT 691
|::| |::| |::| |::| |::|
QY 1 KFLAVWKITYKDT 13

RESULT 15
ID Q84896 PRELIMINARY; PRT; 878 AA.
AC Q84896;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEINS (FRAGMENTS).
GN VP-1.
OS POLIOVIRUS TYPE 3.
OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE;
OC ENTEROVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAUKETT H;
RX MEDLINE; 94160574.
RA HUOVILAINEN A., KINNUNEN L., POYRY T., LAAKSONEN L., ROIVAINEN M.,
RA HOVI T.;
RT "Poliovirus type 3/Saukett: antigenic and structural correlates of
RT sequence variation in the capsid proteins.";
RL VIROLOGY 199:228-232(1994).
DR EMBL; L23848; G388337; -.
DR PFAM; PF00073; rhv; 3.
FT CHAIN 1 >69 CAPSID PROTEIN.
FT NON_CONS 59 70
FT CHAIN 70 >340 CAPSID PROTEIN.
FT NON_CONS 340 341
FT CHAIN 341 >578 CAPSID PROTEIN.
FT NON_CONS 578 579
FT CHAIN 579 >878 CAPSID PROTEIN.
FT NON_TER 878 878
SQ SEQUENCE 878 AA; 97326 MW; 421113F9 CRC32;

Query Match 74.0%; Score 77; DB 14; Length 878;
Best Local Similarity 59.2%; Pred. No. 2.63e-03;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 679 KLEATWRITYKDT 691
|::| |::| |::| |::| |::|
QY 1 KFLAVWKITYKDT 13

Search completed: Wed Aug 4 15:31:33 1999
Job time : 10 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:32:11 1999; MasPar time 9.05 Seconds
70.848 Million cell updates/sec
Tabular output not generated.

Title: >US-09-049-847-4
Description: (1-13) from US09049847.ppe
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 403756 seqs, 49297578 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1:P9 2:U60 3:U7 4:U80 5:U81 6:U82 7:U83 8:U84A 9:U84B
10:U85 11:U86 12:U87 13:U88 14:U89 15:U90 16:U91 17:U92
18:U93 19:NEWU8 20:NEWU9

Statistics: Mean 20.143; Variance 65.088; scale 0.309

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	104	100.0	13	15	US-09-049-	Sequence 4, Applicatio	2.16e-03
2	87	83.7	13	7	US-08-395-	Sequence 5, Applicatio	1.69e-01
3	87	83.7	13	10	US-08-575-	Sequence 5, Applicatio	1.69e-01
4	87	83.7	13	9	US-08-468-	Sequence 3, Applicatio	1.69e-01
5	87	83.7	13	15	US-09-076-	Sequence 3, Applicatio	1.69e-01
6	79	76.0	2206	9	US-08-465-	Sequence 2, Applicatio	1.24e+00
7	70	67.3	322	12	US-08-701-	Sequence 11, Applicati	1.10e+01
8	70	67.3	322	16	US-09-188-	Sequence 11, Applicati	1.10e+01
9	62	59.6	76	15	US-09-134-	Sequence 6406, Applic	7.21e+01
10	62	59.6	111	1	PCT-US97-0	Sequence 375, Applicat	7.21e+01
11	62	59.6	111	13	US-08-858-	Sequence 375, Applicat	7.21e+01
12	62	59.6	606	16	US-09-107-	Sequence 4683, Applic	7.21e+01
13	61	58.7	80	16	US-09-134-	Sequence 5353, Applic	9.07e+01
14	61	58.7	642	17	US-09-253-	Sequence 25, Applicati	9.07e+01
15	61	58.7	642	13	US-08-812-	Sequence 21, Applicati	9.07e+01
16	61	58.7	642	14	US-08-926-	Sequence 25, Applicati	9.07e+01
17	61	58.7	871	16	US-09-107-	Sequence 6227, Applic	9.07e+01
18	59	56.7	300	12	US-08-727-	Sequence 11, Applicati	1.43e+02
19	59	56.7	315	15	US-09-030-	Sequence 112, Applicat	1.43e+02
20	59	56.7	315	15	US-09-030-	Sequence 112, Applicat	1.43e+02
21	59	56.7	315	17	US-09-288-	Sequence 112, Applicat	1.43e+02

22	59	56.7	315	15	US-09-020-	Sequence 112, Applicat	1.43e+02
23	59	56.7	315	15	US-09-020-	Sequence 112, Applicat	1.43e+02
24	59	56.7	410	16	US-09-134-	Sequence 4253, Applic	1.43e+02
25	59	56.7	665	14	US-08-946-	Sequence 25, Applicati	1.43e+02
26	59	56.7	746	17	US-09-248-	Sequence 19979, Applic	1.43e+02
27	59	56.7	746	2	US-50-096-	Sequence 19979, Applic	1.43e+02
28	58	55.8	422	17	US-09-248-	Sequence 16762, Applic	1.79e+02
29	58	55.8	422	2	US-50-096-	Sequence 16762, Applic	1.79e+02
30	58	55.8	552	16	US-09-134-	Sequence 3838, Applic	1.79e+02
31	58	55.8	847	13	US-08-881-	Sequence 164, Applicat	1.79e+02
32	58	55.8	847	1	PCT-US98-0	Sequence 534, Applicat	1.79e+02
33	58	55.8	876	12	US-08-785-	Sequence 2, Applicatio	1.79e+02
34	58	55.8	876	14	US-08-913-	Sequence 2, Applicatio	1.79e+02
35	58	55.8	876	15	US-09-012-	Sequence 2, Applicatio	1.79e+02
36	58	55.8	877	13	US-08-827-	Sequence 4022, Applic	1.79e+02
37	58	55.8	2052	15	US-09-045-	Sequence 2, Applicatio	1.79e+02
38	58	55.8	2052	15	US-09-045-	Sequence 2, Applicatio	1.79e+02
39	57	54.8	434	14	US-08-973-	Sequence 4, Applicatio	2.24e+02
40	57	54.8	434	10	US-08-529-	Sequence 2, Applicatio	2.24e+02
41	57	54.8	434	10	US-08-529-	Sequence 2, Applicatio	2.24e+02
42	57	54.8	434	16	US-09-122-	Sequence 2, Applicatio	2.24e+02
43	57	54.8	434	10	US-08-529-	Sequence 2, Applicatio	2.24e+02
44	56	53.8	399	16	US-09-107-	Sequence 4714, Applic	2.81e+02
45	56	53.8	1886	1	PCT-US97-1	Sequence 3, Applicatio	2.81e+02

ALIGNMENTS

RESULT 1
ID US-09-049-847-4 STANDARD; PRT; 13 AA.

XX xxxxxxx

XX

DT

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 4, Application US/09049847

Sequence 4, Application US/09049847

GENERAL INFORMATION:

APPLICANT: Bay, Sylvie

APPLICANT: Cantacuzene, Daniele

APPLICANT: Leclerc, Claude

APPLICANT: Lo-Man, Richard

TITLE OF INVENTION: Multiple antigen alycopeptide carbohydrate, vaccine

TITLE OF INVENTION: comprising the same and use thereof

FILE REFERENCE: 102.166A

CURRENT APPLICATION NUMBER: US/09/049,847

CURRENT FILING DATE: 1998-03-27

EARLIER APPLICATION NUMBER: 60/041,726

EARLIER FILING DATE: 1997-03-27

NUMBER OF SEQ ID NOS: 4

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4

LENGTH: 13

TYPE: PRT

ORGANISM: Clostridium tetani

SEQUENCE 13 AA; 1613 MW; 1193 CN;

Query Match 100.0%; Score 104; DB 15; Length 13;

Best Local Similarity 100.0%; Pred. No. 2.16e-03;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KFLAVWKITYKDT 13

|||||

QY 1 KFLAVWKITYKDT 13

RESULT 2

ID US-08-395-204-5 STANDARD; PRT; 13 AA.

XX

AC xxxxxx

XX

DT

XX Sequence 5, Application US/08395204
DE Sequence 5, Application US/08395204
XX GENERAL INFORMATION:
CC APPLICANT: Ladtant, Daniel
CC APPLICANT: Leclerc, Claude
CC APPLICANT: Sebo, Peter
CC APPLICANT: Ullmann, Agnes
CC TITLE OF INVENTION: Recombinant Mutants for Inducing
CC TITLE OF INVENTION: Specific Immune Responses
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/395,204
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/871,795
CC FILING DATE: 21-APR-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 03495-0109-00000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1613 MW; 1196 CN;
Query Match 83.7%; Score 87; DB 7; Length 13;
Best Local Similarity 84.6%; Pred. No. 1.69e-01;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Db 1 KLEAVWKITYKDT 13
QY 1 KFLAVWKITYKDI 13
RESULT 3
ID US-08-575-879-5 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 5, Application US/08575879
XX Sequence 5, Application US/08575879
CC GENERAL INFORMATION:
CC APPLICANT: Fayolle, Catherine
CC APPLICANT: Ladtant, Daniel
CC APPLICANT: Leclerc, Claude
CC APPLICANT: Ullmann, Agnes
CC TITLE OF INVENTION: Recombinant Adenylate Cyclase Toxin For
CC TITLE OF INVENTION: Inducing Specific Immune Responses

CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner, L.L.P.
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: DC
CC COUNTRY: USA
CC ZIP: 20005-3315
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/575,879
CC FILING DATE: 22-DEC-1995
CC CLASSIFICATION: 424
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 03495-0109-05000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 13 AA; 1613 MW; 1196 CN;
Query Match 83.7%; Score 87; DB 10; Length 13;
Best Local Similarity 84.6%; Pred. No. 1.69e-01;
Matches 11; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Db 1 KLEAVWKITYKDT 13
QY 1 KFLAVWKITYKDT 13
RESULT 4
ID US-08-468-021-3 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08468021
XX Sequence 3, Application US/08468021
CC GENERAL INFORMATION:
CC APPLICANT: GENGOUX, CHRISTINE; LeCLERC, CLAUDE
CC TITLE OF INVENTION: ANTIGEN-CARRYING
CC TITLE OF INVENTION: MICROPARTICLES AND THEIR USE IN THE IN THE
CC TITLE OF INVENTION: INDUCTION OF HUMORAL OR CELLULAR RESPONSES
CC NUMBER OF SEQUENCES: 5
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BIERMAN & MUSERLIAN
CC STREET: 600 THIRD AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: U.S.A.
CC ZIP: 10016
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: ASC II
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,021

CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/397,286
CC FILING DATE: 10-MAR-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/FR93/00876
CC FILING DATE: 13-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: FR/92/10879
CC FILING DATE: 11-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: CHARLES A. MUSERLIAN
CC REGISTRATION NUMBER: 19,683
CC REFERENCE/DOCKET NUMBER: 102.162-CON
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-661-8000
CC TELEFAX: 212-661-8002
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 13
CC TYPE: AMINO ACID
CC STRANDEDNESS: UNKNOWN
CC TOPOLOGY: UNKNOWN
CC MOLECULE TYPE: PEPTIDE
CC SEQUENCE 13 AA; 1613 MW; 1196 CN;

Query Match 83.7%; Score 87; DB 9; Length 13;
Best Local Similarity 84.6%; Pred. No. 1.69e-01;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 KLFVAVWKITYKDT 13
QY 1 KFLAVWKITYKDT 13

RESULT 5
ID US-09-076-646-3 STANDARD; PRT; 13 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX

Sequence 3, Application US/09076646
Sequence 3, Application US/09076646
GENERAL INFORMATION:
APPLICANT: GENGOUX, CHRISTINE
APPLICANT: LECLERC, CLAUDE
TITLE OF INVENTION: ANTIGEN-CARRYING MICROPARTICLES AND THEIR USE IN THE
TITLE OF INVENTION: INDUCTION OF HUMORAL OR CELLULAR RESPONSES
FILE REFERENCE: 102.162-1
CURRENT APPLICATION NUMBER: US/09/076,646
CURRENT FILING DATE: 1998-05-12
EARLIER APPLICATION NUMBER: US 08/397,286
EARLIER FILING DATE: 1995-04-28
EARLIER APPLICATION NUMBER: PCT/FR93/00876
EARLIER FILING DATE: 1993-09-13
EARLIER APPLICATION NUMBER: FR 9210879
EARLIER FILING DATE: 1992-09-11
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 13
TYPE: PRT
ORGANISM: POLIOMYELITIS VIRUS
SEQUENCE 13 AA; 1613 MW; 1196 CN;

Query Match 83.7%; Score 87; DB 15; Length 13;
Best Local Similarity 84.6%; Pred. No. 1.69e-01;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 KLFVAVWKITYKDT 13

QY 1 KFLAVWKITYKDT 13
RESULT 6
ID US-08-465-250-2 STANDARD; PRT; 2206 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 2, Application US/08465250
Sequence 2, Application US/08465250
GENERAL INFORMATION:
APPLICANT: Racaniello, Vincent
APPLICANT: Tateam, Joanne M.
APPLICANT: Weeks-Levy, Carolyn L.
TITLE OF INVENTION: METHODS FOR PRODUCING RNA VIRUSES FROM
TITLE OF INVENTION: CDNA
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release 1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,250
FILING DATE: 6-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 36607-E-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2206 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 2206 AA; 246209 MW; 25431833 CN;
Query Match 76.0%; Score 79; DB 9; Length 2206;
Best Local Similarity 69.2%; Pred. No. 1.24e+00;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
Db 679 KLFAMWRITYKDT 691
QY 1 KFLAVWKITYKDT 13
RESULT 7
ID US-08-701-191A-11 STANDARD; PRT; 322 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 11, Application US/08701191A
Sequence 11, Application US/08701191A
GENERAL INFORMATION:

CC APPLICANT: Moosa Mohammadi, Joseph Schlessinger,
CC APPLICANT: and Stevan R. Hubbard
CC TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
CC TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE
CC NUMBER OF SEQUENCES: 41
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lyon & Lyon
CC STREET: 633 West Fifth Street
CC SIREET: Suite 4700
CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90071-2066
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC MEDIUM TYPE: storage
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: IBM P.C. DOS 5.0
CC SOFTWARE: FastSEQ for Windows 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/701,191A
CC FILING DATE: August 21, 1996
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warburg, Richard J.
CC REGISTRATION NUMBER: 32,327
CC REFERENCE/DOCKET NUMBER: 227/088
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (213) 489-1600
CC TELEFAX: (213) 955-0440
CC TELEX: 67-3510
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 322 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 322 AA; 37050 MW; 551831 CN;

Query Match 67.3%; Score 70; DB 12; Length 322;
Best Local Similarity 66.7%; Pred. No. 1.10e+01;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 34 FGEVWKATYKET 45
| ||| |||||
QY 2 FLAVWKITYKDT 13

RESULT 8
ID US-09-188-809-11 STANDARD; PRI; 322 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 11, Application US/09188809
XX
CC Sequence 11, Application US/09188809
CC GENERAL INFORMATION:
CC APPLICANT: Moosa Mohammadi
CC APPLICANT: Joseph Schlessinger
CC TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
CC TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE
CC NUMBER OF SEQUENCES: 41
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Lyon & Lyon
CC STREET: 633 West Fifth Street
CC STREET: Suite 4700

CC CITY: Los Angeles
CC STATE: California
CC COUNTRY: U.S.A.
CC ZIP: 90071-2066
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC MEDIUM TYPE: storage
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: IBM P.C. DOS 5.0
CC SOFTWARE: FastSEQ for Windows 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/188,809
CC FILING DATE: November 9, 1998
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/701,191
CC FILING DATE: August 21, 1996
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Warburg, Richard J.
CC REGISTRATION NUMBER: 32,327
CC REFERENCE/DOCKET NUMBER: 238/091
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (213) 489-1600
CC TELEFAX: (213) 955-0440
CC TELEX: 67-3510
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 322 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 322 AA; 37050 MW; 551831 CN;

Query Match 67.3%; Score 70; DB 16; Length 322;
Best Local Similarity 66.7%; Pred. No. 1.10e+01;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Db 34 FGEVWKATYKET 45
| ||| |||||
QY 2 FLAVWKITYKDT 13

RESULT 9
ID US-09-134-000-6406 STANDARD; PRI; 76 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 6406, Application US/09134000A
XX
CC Sequence 6406, Application US/09134000A
CC GENERAL INFORMATION:

CC APPLICANT: Lynn Doucette-Stamm et al
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCO
CC TITLE OF INVENTION: FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
CC FILE REFERENCE: GTC-005
CC CURRENT APPLICATION NUMBER: US/09/134,000A
CC CURRENT FILING DATE: 1998-08-13
CC NUMBER OF SEQ ID NOS: 6810
CC SEQ ID NO 6406
CC LENGTH: 76
CC TYPE: PRT
CC ORGANISM: Enterococcus faecalis
CC SEQUENCE 76 AA; 8502 MW; 28938 CN;

Query Match 59.6%; Score 62; DB 16; Length 76;
Best Local Similarity 58.3%; Pred. No. 7.21e+01;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 5 KFVAVWQIQIKE 16


```
QY      ||:|||| | ||:
1 KFLAVWKITYKD 12

RESULT 10
ID PCT-US97-07950-375 STANDARD; PRT; 111 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 375, Application PC/TUS9707950
XX
CC Sequence 375, Application PC/TUS9707950
CC GENERAL INFORMATION:
CC APPLICANT: Black, Michael
CC APPLICANT: Hodgson, John
CC APPLICANT: Knowles, David
CC APPLICANT: Nicholas, Richard
CC APPLICANT: Stodola, Robert
CC TITLE OF INVENTION: Novel Compounds
CC NUMBER OF SEQUENCES: 552
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCI/US97/07950
CC FILING DATE: 14-MAY-1997
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/017670
CC FILING DATE: 14-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gimmi, Edward R
CC REGISTRATION NUMBER: 38,891
CC REFERENCE/DOCKET NUMBER: P50475
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-4478
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 375:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 111 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: Protein
CC SEQUENCE 111 AA; 13251 MW; 68279 CN;

Query Match 59.6%; Score 62; DB 1; Length 111;
Best Local Similarity 60.0%; Pred. No. 7.21e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 66 KYLSAWVITY 75
|:|:| | | |
QY 1 KFLAVWKITY 10

RESULT 11
ID US-08-858-207A-375 STANDARD; PRT; 111 AA.
XX
AC xxxxxx
XX
DT
```

```
XX Sequence 375, Application US/08858207A
DE
XX
CC Sequence 375, Application US/08858207A
CC GENERAL INFORMATION:
CC APPLICANT: Black, Michael
CC APPLICANT: Hodgson, John
CC APPLICANT: Knowles, David
CC APPLICANT: Nicholas, Richard
CC APPLICANT: Stodola, Robert
CC TITLE OF INVENTION: Novel Compounds
CC NUMBER OF SEQUENCES: 552
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: SmithKline Beecham Corporation
CC STREET: 709 Swedeland Road
CC CITY: King of Prussia
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19406-0939
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/858,207A
CC FILING DATE: 09-MAY-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/017670
CC FILING DATE: 14-MAY-1996
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Gimmi, Edward R
CC REGISTRATION NUMBER: 38,891
CC REFERENCE/DOCKET NUMBER: P50475
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-270-4478
CC TELEFAX: 610-270-5090
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 375:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 111 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: None
CC SEQUENCE 111 AA; 13251 MW; 68279 CN;

Query Match 59.6%; Score 62; DB 13; Length 111;
Best Local Similarity 60.0%; Pred. No. 7.21e+01;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 66 KYLSAWVITY 75
|:|:| | | |
QY 1 KFLAVWKITY 10

RESULT 12
ID US-09-107-532-4683 STANDARD; PRT; 606 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 4683, Application US/09107532
XX
CC Sequence 4683, Application US/09107532
CC GENERAL INFORMATION:
CC APPLICANT: Lynn A Doucette-Stamm and David Bush
CC TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
CC TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR
CC TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
CC NUMBER OF SEQUENCES: 7308
```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: GENOME THERAPEUTICS CORPORATION
CC STREET: 100 Beaver Street
CC CITY: Waltham
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02354
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: CD-ROM ISO9660
CC OPERATING SYSTEM:
CC SOFTWARE:
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/107,532
CC FILING DATE:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/ 085598
CC FILING DATE: May 14, 1998
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 60/051571
CC FILING DATE: July 2, 1997
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ariniello, Pamela Deneke
CC REGISTRATION NUMBER: 40,489
CC REFERENCE/DOCKET NUMBER: GTC-012
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (781)893-5007
CC TELEFAX: (781)893-8277
CC INFORMATION FOR SEQ ID NO: 4683:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 606 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: YES
CC ORIGINAL SOURCE:
CC ORGANISM: Enterococcus faecium
CC FEATURE:
CC NAME/KEY: misc_feature
CC LOCATION: 1...606
CC SEQUENCE 506 AA; 70613 MW; 2071988 CN;

Query Match 59.6%; Score 62; DB 16; Length 606;
Best Local Similarity 54.5%; Pred. No. 7.21e+01;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 255 FLTNWKIPFED 265
||: ||| : |
QY 2 FLAVWKITYKD 12

RESULT 13
ID US-09-134-000-5353 STANDARD; PRT; 80 AA.
XX xxxxxx

Sequence 5353, Application US/09134000A
Sequence 5353, Application US/09134000A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
FILE REFERENCE: GTC-005

CC CURRENT APPLICATION NUMBER: US/09/134,000A
CC CURRENT FILING DATE: 1998-08-13
CC NUMBER OF SEQ ID NOS: 6810
CC SEQ ID NO 5353
CC LENGTH: 80
CC TYPE: PRT
CC ORGANISM: Enterococcus faecalis

SQ SEQUENCE 80 AA; 9450 MW; 34784 CN;
Query Match 58.7%; Score 61; DB 16; Length 80;
Best Local Similarity 63.6%; Pred. No. 9.07e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Db 41 KFLNIRKRTYK 51
||| :| |||
QY 1 KFLAVWKITYK 11

RESULT 14
ID US-09-253-682-25 STANDARD; PRT; 642 AA.
XX xxxxxx

Sequence 25, Application US/09253682
Sequence 25, Application US/09253682
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
APPLICANT: Cha, Tai-An
TITLE OF INVENTION: NOVEL HUMAN CYTOMEGALOVIRUS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Luann Cserr Attorney at Law
STREET: 750 Arimo Avenue
CITY: Oakland
STATE: CA
COUNTRY: USA
ZIP: 94610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/253,682
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/926,922
FILING DATE: September 10, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Cserr, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR 11A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-834-1448
TELEFAX: 510-839-7810
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 642 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: tol.21
FEATURE:
NAME/KEY: Protein
LOCATION: 1..642
OTHER INFORMATION: /label= UL150
SQ SEQUENCE 642 AA; 70848 MW; 2000477 CN;

Query Match 58.7%; Score 61; DB 17; Length 642;
Best Local Similarity 63.6%; Pred. No. 9.07e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQRTYNDT 414
||: ||| |||
QY 3 LAVWKITYKDT 13

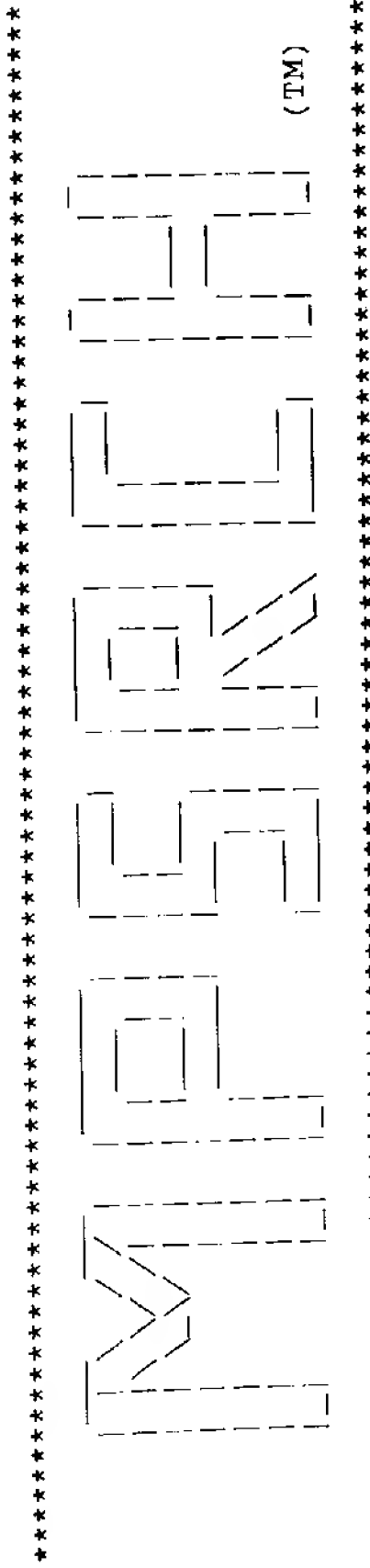
```
RESULT 15
ID US-08-812-716-21 STANDARD; PRT; 642 AA.
XX
AC xxxxxx
XX
DI
XX
DE
XX
Sequence 21, Application US/08812716
Sequence 21, Application US/08812716
GENERAL INFORMATION:
APPLICANT: KEMBLE, George
APPLICANT: DUKE, Gregory
APPLICANT: SPATE, Richard
TITLE OF INVENTION: ATTENUATION OF CYTOMEGALOVIRUS
TITLE OF INVENTION: VIRULENCE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: AVIRON
STREET: 297 N. Bernardo Avenue
CITY: Mountain View
STATE: California
COUNTRY: USA
ZIP: 94043
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/812,716
FILING DATE: 06-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENCY INFORMATION:
NAME: DUNN, Tracy
REGISTRATION NUMBER: 34,587
REFERENCE/DOCKET NUMBER: AVIR-5032
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650.919.6637
TELEFAX: 650.919.6610
TELEX:
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 642 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SQ SEQUENCE 642 AA; 70848 MW; 2000477 CN;

Query Match 58.7%; Score 61; DB 13; Length 642;
Best Local Similarity 63.6%; Pred. No. 9.07e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWORTYNDT 414
||:| ||:|
QY 3 LAVWKITYKDT 13

Search completed: Wed Aug 4 15:32:24 1999
Job time : 13 secs.
```

This Page Blank (uspio)



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Aug 4 15:30:02 1999; MasPar time 9.41 Seconds
Tabular output not generated. 29.384 Million cell updates/sec

Title: >US-09-049-847-4
Description: (1-13) from US09049847.pap
Perfect Score: 104
Sequence: 1 KFLAVWKITYKDT 13

Scoring table: PAM 150
Gap 15

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 19.550; Variance 63.096; scale 0.310

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description	Pred. No.
1	104	100.0	13 36	W67036	Poliovirus antigen.	6.52e-04
2	87	83.7	13 36	W67037	Poliovirus antigen.	5.31e-02
3	87	83.7	106 1	P90493	Poliovirus VP-1 capsid	5.31e-02
4	87	83.7	106 4	P40106	Sequence of a peptide	5.31e-02
5	87	83.7	2209 5	P20037	Sequence encoded by a	5.31e-02
6	79	76.0	2206 4	R22210	True type 3 poliovirus	3.97e-01
7	62	59.6	111 34	W38579	Streptococcus pneumoniae	2.41e+01
8	61	58.7	500 38	W83330	Thermus flavus amylo	3.04e+01
9	61	58.7	642 19	W05519	HCMV Toledo strain UL	3.04e+01
10	59	56.7	315 35	W69384	Prostate tumour speci	4.83e+01
11	59	56.7	315 36	W71868	Amino acid encoded by	4.83e+01
12	59	56.7	665 32	W54425	Human PS112 protein s	4.83e+01
13	58	55.8	876 23	W21898	Alanyl-tRNA synthetas	6.07e+01
14	57	54.8	151 30	W40495	Human heart caveolin	7.62e+01
15	57	54.8	434 20	W01520	MagA protein.	7.62e+01
16	57	54.8	434 28	W27454	Magnetospirillum sp.	7.62e+01

17	56	53.8	1886 31	W54241	Rattus norvegicus mut	9.56e+01
18	55	52.9	1096 1	P82507	pullulanase protein.	1.20e+02
19	54	51.9	14 30	W53471	P2 predominant PNS my	1.50e+02
20	54	51.9	136 30	W40228	Bovine myelin P2 prot	1.50e+02
21	54	51.9	136 30	W40227	Human myelin P2 prote	1.50e+02
22	53	51.0	253 4	R22323	Marek Disease Virus U	1.87e+02
23	53	51.0	355 21	W13105	Marek's disease virus	1.87e+02
24	53	51.0	355 21	W11475	Marek's disease virus	1.87e+02
25	53	51.0	466 22	W09825	UDP-glucose:thiohydro	1.87e+02
26	53	51.0	634 38	W89888	Antigen from cluster	1.87e+02
27	53	51.0	637 20	W11704	High affinity Na+dep	1.87e+02
28	53	51.0	667 31	W57224	Rat proline transport	1.87e+02
29	53	51.0	842 4	P93712	Sequence of infectiou	1.87e+02
30	52	50.0	391 8	R43273	Peptide encoded by La	2.33e+02
31	52	50.0	559 2	R06518	Brassica microspore-s	2.33e+02
32	52	50.0	559 20	W08380	Brassica napus micros	2.33e+02
33	51	49.0	401 7	R39386	Gsea enzyme.	2.90e+02
34	51	49.0	516 39	W67617	P. chrysogenum phenyl	2.90e+02
35	51	49.0	968 28	W41309	CF-5 pathogen resista	2.90e+02
36	50	48.1	538 20	W10058	Human wild-type bile	3.60e+02
37	50	48.1	745 19	R99257	Human bile salt-stimu	3.60e+02
38	50	48.1	1435 13	R70232	P. falciparum SABP.	3.60e+02
39	50	48.1	2723 31	W56448	Fragment HGJ1775 of a	3.60e+02
40	50	48.1	2873 39	W89452	Hepatitis G virus PNF	3.60e+02
41	50	48.1	2873 31	W56441	Fragment HGJ606 of a	3.60e+02
42	50	48.1	2873 18	R90796	HGV-PNF 2161 polyprot	3.60e+02
43	50	48.1	2910 39	W89458	Hepatitis G virus var	3.60e+02
44	50	48.1	2969 31	W56442	Fragment HGJ1737 of a	3.60e+02
45	50	48.1	2969 31	W56443	Fragment HGJ1741 of a	3.60e+02

ALIGNMENTS

RESULT 1

ID W67036 standard; peptide; 13 AA.

AC W67036;

DT 15-DEC-1998 (first entry)

DE Poliovirus antigen.

KW Tetanus toxin; vaccine; antibody; carbohydrate peptide conjugate;

KW dendrimeric poly-lysine; epitope; tumour.

OS Poliovirus.

PN WQ9843677-A1.

PD 08-OCT-1998.

PF 27-MAR-1998; E01922.

PR 27-MAR-1997; US-041726.

PA (INSP) INST PASTEUR.

PI Bay S, Cantacuzene D, Leclerc C, Lo-man R;

DR WPI; 98-557071/47.

PT Carbohydrate peptide conjugate used as vaccine - comprises carrier

PT with dendrimeric poly-lysine enabling multiple epitopes to be

PT covalently attached

PS Disclosure; Page 14; 55pp; English.

CC The invention relates to a new carbohydrate peptide conjugate, which

CC comprises a carrier with a dendrimeric poly-lysine enabling multiple

CC epitopes to be covalently attached to it. Also claimed are: (1) an

CC antibody purified from biological fluid or cells of organisms

CC administered with the carbohydrate peptide conjugate, and (2) a diagnosis

CC kit comprising antigen-specific antibodies elicited by immunisation with

CC the carbohydrate peptide conjugate. The peptide conjugate, antibody and

CC diagnosis kit are used to provide pharmaceutical compositions and

CC vaccines against tumours. These can be used to support an immune response

CC against viral infections caused by hepatitis virus, HIV or cytomegalo

CC virus. They can be used to enhance immune responses, especially B- and T-

CC cell responses, of humans and animals against bacterial infections. The

CC carbohydrate peptide conjugate stimulates the antibody and T-cell

CC response without stimulating undesired immune responses. The composition

CC is capable of increasing the survival of tumour bearing humans and

CC animals. The present sequence represents a poliovirus antigen which

CC forms part of a carbohydrate peptide conjugate.

SQ Sequence 13 AA;

Query Match 100.0%; Score 104; DB 36; Length 13;

Best Local Similarity 100.0%; Pred. No. 6.52e-04;

FT /label= 3b
 FT 1031..1127
 FT /label= 5b
 FT 1128..1456
 FT /label= x
 FT 1457..1543
 FT /label= 1b
 FT 1544..1565
 FT /label= VPg
 FT 1566..1748
 FT /label= 2
 FT 1749..2209
 FT /label= 4(p(63))
 PN W08203632-A.
 PD 28-OCT-1982.
 PD 12-NOV-1981; 320525.
 PF 20-APR-1981; US-255879.
 PR 12-NOV-1981; US-320525.
 PA (MASI) MASSACHUSETTS INST TECH.
 PI Baltimore D; Racaniello VR;
 DR WPI; 82-95059E/44 (95059E).
 DR N-PSDB; N20042.
 DR Prodn. of CDNA representing viral RNA sequences - by
 PT transcription, insertion into vector and host cell transformation
 PS Example; Table 1, pages 25-31; 50pp; English.
 CC Plasmid pVR106 was produced by combining plasmids pVR104 and pVR105.
 CC It contains a full-length CDNA copy of the poliovirus genome. E.coli
 CC HB101 contg. this plasmid has been registered as ATCC 31844. The
 CC full-length poliovirus CDNA molecule is itself infectious and can be
 CC introduced into cells and these cultured to produce RNA virus.
 CC Alternatively, the infectious CDNA can be treated with mutagens and
 CC the altered material used to infect cells so that attenuated viral
 CC RNA is prod. and this used to make vaccines. For antibody prodn.,
 CC CDNA capable of directing antigen prodn. is selected and isolated and
 CC incorporated into cells which are incubated to produce RNA antigen.
 SQ Sequence 2209 AA;

Query Match 83.7%; Score 87; DB 5; Length 2209;
 Best Local Similarity 84.6%; Pred. No. 5.31e-02;
 Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 682 klfavwkyktdt 694
 QY 1 KFLAVWKITYKDT 13

RESULT 6
 ID R22210 standard; Protein; 2206 AA.
 AC R22210;
 DI 14-JUL-1992 (first entry)
 DE True type 3 poliovirus protein from LED3.
 KW RNA virus; error reduction.
 OS Poliovirus.
 PN W09203538-A.
 PD 05-MAR-1992.
 PF 20-AUG-1991; J05890.
 PR 20-AUG-1990; US-570000.
 PR 20-AUG-1990; US-559916.
 PA (UINY-) COLUMBIA UNIV NEW Y.
 PI Racaniello V, Tatem JM, Weekslevy CL;
 DR WPI; 92-096882/12.
 DR N-PSDB; Q22965.
 DR New vaccine against infectious polio-virus comprises RNA virus -
 PT for producing RNA virus CDNA and viable RNA virus
 PS Disclosure; Fig 6; 110pp; English.
 CC The protein sequence was deduced from the cDNA sequence of p3
 CC poliovirus obtd as in Q22965. The cDNA sequence is that of a
 CC true RNA virus, i.e. the CDNA directs the prodn. of a viable
 CC RNA virus which is phenotypically similar to the source virus.
 CC The full length CDNA in pLED3 was infectious. In vitro
 CC transcription of pLED3 CDNA using T7 RNA polymerase produced
 CC RNAs which possessed several erroneous amino acids. The RNA
 CC viruses are used in vaccines against polio. The screening method

CC can be used during amplification of the source virus for vaccine
 CC prodn. to ensure maintenance of C at position 2493 in the viral
 CC genome i.e. increasing the attenuation. The new prod. overcomes
 CC the problem of errors introduced during replication of ss RNA,
 CC which is much higher than for ds DNA.
 SQ Sequence 2206 AA;

Query Match 76.0%; Score 79; DB 4; Length 2206;
 Best Local Similarity 69.2%; Pred. No. 3.97e-01;
 Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Db 679 klfamwritykdtd 691
 QY 1 KFLAVWKITYKDT 13

RESULT 7
 ID W38579 standard; Protein; 111 AA.
 AC W38579;
 DT 06-NOV-1998 (first entry)
 DE Streptococcus pneumoniae protein of unknown function.
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis.
 OS Streptococcus pneumoniae.
 PN W09743303-A1.
 PD 20-NOV-1997.
 PF 14-MAY-1997; U07950.
 PR 14-MAY-1996; US-017670.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Stodola RK;
 DR WPI; 98-008793/01.
 DR N-PSDB; T986730.
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PI diagnosing anti-microbial agents for treatment of bacterial
 PI infections
 PS Claim 12; Page 350; 483pp; English.
 CC This sequence represents a Streptococcus pneumoniae protein of
 CC unknown function, and is encoded by a DNA sequence of the invention.
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 SQ Sequence 111 AA;

Query Match 59.6%; Score 62; DB 34; Length 111;
 Best Local Similarity 60.0%; Pred. No. 2.41e+01;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 66 kylsawvity 75
 QY 1 KFLAVWKITY 10

RESULT 8
 ID W83330 standard; Protein; 500 AA.
 AC W83330;

DT 01-MAR-1999 (first entry)
DE Thermus flavus amylomaltase.
KW Thermus flavus; amylomaltase; heat resistant; cyclic glucan;
KW intramolecular transglycosylation; alpha-glucan; food.
OS Thermus flavus.
PN EP-884384-A2.
PD 16-DEC-1998.
PF 13-MAY-1998; 250162.
PR 07-MAY-1998; JP-125121.
PR 13-MAY-1997; JP-122635.
PA (EZAKI) EZAKI GLICO CO LTD.
PI Fujii K, Okada S, Takaha T, Takata H, Terada Y,
PI Yanase M;
DR WPI; 99-026580/03.
DR N-PSDB; V72539.
PT New amylomaltase catalyses intra-molecular trans-glycosylation of
PT alpha-glucans - used as additive to e.g. rice products, snacks,
PT wheat products, noodles, processed seafood, frozen or refrigerated,
PT foods, baby foods or drinks
PS Claim 3; Page 17-18; 32pp; English.
CC The present sequence is an amylomaltase isolated from Thermus flavus.
CC Amylomaltase catalyses intramolecular transglycosylation of alpha-glucans
CC to generate cyclic glucans, has no hydrolase activity, has an optimum
CC temperature of 65-70 degrees Celsius, remains active at 60 degrees
CC Celsius for at least 10 minutes, is inactivated at 100 degrees Celsius
CC within 15 minutes and has an optimum pH of 5.5. Amylomaltase can be used
CC to produce a cyclic glucan by cyclising an alpha-glucan and collecting
CC and purifying the cyclic glucan (especially where the cyclic glucan
CC comprises a cyclic alpha-1,4-glucan, a branched cyclic glucan and/or a
CC branching enzyme is also used in the cyclisation step). The amylomaltase
CC is used to produce food by adding it to a food material before or
CC immediately after cooking so that the amylomaltase acts on starch in the
CC food material to produce a cyclic glucan (especially where the food is
CC selected from rice products, Japanese desserts, snacks, wheat products,
CC noodles, gyoza skins, shumai skins, processed seafoods, frozen or
CC refrigerated processed foods, weaning foods, baby foods, pet foods,
CC animal feeds, drinks, sports foods and nutritional supplements).
SQ Sequence 500 AA;

Query Match 58.7%; Score 61; DB 38; Length 500;
Best Local Similarity 41.7%; Pred. No. 3.04e+01;
Matches 5; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Db 416 ryladwgitf 427
QY 1 KFLAVWKITYKD 12

RESULT 9
ID W05519 standard; Protein; 542 AA.
AC W05519;
DT 15-JAN-1997 (first entry)
DE HCMV Toledo strain UL150 protein (clone tol.21).
KW CMV; HCMV; vaccine; diagnosis; UL150.
OS Human cytomegalovirus Toledo strain.
PN W09630387-A1.
PD 03-OCT-1996.
PF 26-MAR-1996; U04100.
PR 31-MAR-1995; US-414926.
PA (AVIR-) AVIRON.
PI Cha T, Spaete R;
DR WPI; 96-455265/45.
DR N-PSDB; T41418.
PT New isolated human cytomegalovirus nucleic acid - from Towne and
PT Toledo strains, used to develop prods. for the diagnosis, prevention
PT and treatment of human CMV infections
PS Claim 5; Page 85-88; 150pp; English.
CC Novel protein UL150 (W05519) is the product of an open reading
CC frame found in a novel nucleic acid (T41418) isolated from the
CC Toledo strain of human cytomegalovirus (HCMV). UL150 and other
CC novel (see also W05502-20) and known (see also W05500-01) proteins
CC of the Toledo strain, as well as new proteins (see also W05521-24)
CC from HCMV Towne, can be produced in transformed host cells and used

CC in the prodn. of subunit vaccines against HCMV. They may be
CC surface glycoproteins that are immunogenic or responsible for
CC tissue tropism, or may influence the immune response of an infected
CC individual.
SQ Sequence 642 AA;

Query Match 58.7%; Score 61; DB 19; Length 642;
Best Local Similarity 63.6%; Pred. No. 3.04e+01;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 laiwqrtyndt 414
QY 3 LAVWKITYKDT 13

RESULT 10
ID W69384 standard; Protein; 315 AA.
AC W69384;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone J1-17 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
OS Homo sapiens.
PN W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-480805/41.
DR N-PSDB; V58585.

PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 86-87; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 315 AA;

Query Match 56.7%; Score 59; DB 35; Length 315;
Best Local Similarity 55.6%; Pred. No. 4.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 37 flgvwlvay 45
QY 2 FLAVWKITY 10

RESULT 11
ID W71868 standard; Protein; 315 AA.
AC W71868;
DT 06-JAN-1999 (first entry)
DE Amino acid encoded by prostate tumour clone J1-17.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
PN W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI; 98-609886/51.
DR N-PSDB; V61200.

PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Example 1; Page 81-82; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 315 AA;

Query Match 56.7%; Score 59; DB 36; Length 315;
Best Local Similarity 55.6%; Pred. No. 4.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 37 flgvwlvay 45
:::|:::
QY 2 FLAVWKITY 10

RESULT 12
ID W54425 standard; Protein; 665 AA.
AC W54425;
DT 15-SEP-1998 (first entry)
DE Human PS112 protein sequence from gene-specific clones.
KW Prostate; disease; PS112 gene; detection; diagnosis; cancer;
KW treatment; antibody.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Protein 1..665
FT /label= PS112
FT /note= "partial sequence"
PN W09815657-A1.
PD 16-APR-1998.
PE 08-OCT-1997; U18290.
PR 08-OCT-1996; US-727688.
PA (ABBO) ABBOTT LAB.
PI Cohen M, Friedman PN, Gordon J, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI; 98-240838/21.
PT Detecting a target PS112 polynucleotide - used for diagnosing
PT prostate cancer
PS Example 10; Page 89-90; 104pp; English.
CC This sequence represents the human PS112 protein derived from various
CC gene specific clones isolated from a prostate library and is used in a
CC novel method of detecting the presence of a target PS112 polynucleotide
CC in a test sample. The method can also be used to detect mRNA of PS112 in
CC a test sample. The method can be used for diagnosis of prostate cancer,
CC as the presence of PS112 is an indicator of prostate cancer. Antibodies
CC against the polypeptides may be used as markers, or to treat prostate
CC cancer.
SQ Sequence 665 AA;

Query Match 56.7%; Score 59; DB 32; Length 665;
Best Local Similarity 55.6%; Pred. No. 4.83e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 388 flgvwlvay 396
:::|:::
QY 2 FLAVWKITY 10

RESULT 13
ID W21898 standard; Protein; 876 AA.
AC W21898;
DT 11-SEP-1997 (first entry)
DE Alanyl-tRNA synthetase from Staph. aureus.
KW tRNA synthetase; Escherichia coli; immunological response; antibody;
KW bacterial infection; adherence; damaged tissue; wound healing;
KW vaccine; skin; protection.
OS Staphylococcus aureus.
PN EP-785260-A1.
PD 23-JUL-1997.

PF 17-JAN-1997; 300309.
PR 19-JAN-1996; GB-001099.
PR 30-OCT-1996; GB-022617.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Hodgson JE, Lawlor EJ;
DR WPI; 97-365935/34.
DR N-PSDB; T73696.
PT DNA encoding alanyl-tRNA synthetase from Staphylococcus aureus WCUH
PT 29 - useful for protection against bacterial infections
PS Claim 1; Fig 2; 35pp; English.
CC The present sequence represents a novel alanyl-tRNA synthetase protein
CC from Staphylococcus aureus strain WCUH29 (NCIMB 40771), which is related
CC by amino acid sequence homology to Escherichia coli alanyl tRNA
CC synthetase. Vectors comprising the DNA (or polynucleotides having at
CC least 70 % identity to it) can be used for the recombinant production
CC of the enzyme. The enzyme or its related DNA (through gene therapy) is
CC used to induce an immunological response in a mammal to generate
CC antibodies to protect against disease. The antibodies protect against
CC invasion of bacteria, e.g. by blocking adherence of bacteria to damaged
CC tissue, including wounds in skin or connective tissue caused by
CC mechanical, chemical or thermal damage or by implantation of in-dwelling
CC devices, or wounds in the mucous membranes. Antagonists are used to
CC inhibit the enzyme, especially to prevent adhesion of bacteria to
CC mammalian extracellular matrix proteins on in-dwelling devices or to
CC extracellular matrix proteins in wounds, or to block tRNA synthetase
CC protein mediated mammalian cell invasion by, e.g. initiating
CC phosphorylation of mammalian tyrosine kinase. Analysing a sample for
CC the presence of the enzyme (or a polypeptide having at least 70 %
CC identity to it) is used for a diagnostic process.
SQ Sequence 876 AA;

Query Match 55.8%; Score 58; DB 23; Length 876;
Best Local Similarity 25.0%; Pred. No. 6.07e+01;
Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 197 rylevwnlvfse 208
:::|:::
QY 1 KFLAVWKITYKD 12

RESULT 14
ID W40495 standard; Protein; 151 AA.
AC W40495;

DT 14-JUL-1998 (first entry)
DE Human heart caveolin protein.
KW Caveolin; human; heart; prophylaxis; treatment; diabetes; obesity;
KW cancer; arteriosclerosis; muscular dystrophy; inhibitor.

OS Homo sapiens.

PN J10087698-A.

PD 07-APR-1998.

PF 03-JUL-1997; 177496.

PR 09-JUL-1996; JP-179666.

PA (TAKE) TAKEDA CHEM IND LTD.

DR WPI; 98-267126/24.

DR N-PSDB; V11143; V11147.

PT New human heart caveolin protein and related DNA - useful for
PT preventing and treating diabetes, obesity, cancer, arteriosclerosis
PT and muscular dystrophy

PS Claim 1; Fig 1; 25pp; Japanese.

CC This sequence represents a human heart calveolin protein. This protein
CC can be used for prophylaxis and treatment of diabetes, obesity, cancer,
CC arteriosclerosis and muscular dystrophy. Calveolin can also be used for
CC identifying compounds which can promote or inhibit its activity.

SQ Sequence 151 AA;

Query Match 54.8%; Score 57; DB 30; Length 151;
Best Local Similarity 55.6%; Pred. No. 7.62e+01;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Db 54 fdsvwkvsvy 62
:::|:::
QY 2 FLAVWKITY 10


```
RESULT 15
ID W01520 standard; Protein; 434 AA.
AC W01520;
DT 28-FEB-1997 (first entry)
DE MagA protein.
KW magA; magnetic; microbe; protein preparation; organic membrane.
OS Magnetospirillum sp.
FH Key Location/Qualifier
FT region 7..380
ET /label= hydrophobic_region
PN J08228782-A.
PD 10-SEP-1996.
PF 18-SEP-1995; 263487.
PR 16-SEP-1994; JP-248700.
PA (MATS/) MATSUNAGA T.
PA (DENK ) TDK CORP.
DR WPI; 96-459147/46.
DR N-PSDB; T46127.
PT magA gene encoding protein bound to organic membrane - covering fine
PI magnetic particles formed in magnetic microbe AMB1, useful for
PT stable prepn. of proteins
PS Claim 2; Page 9-11; 17pp; Japanese.
CC The magA protein binds to an organic membrane covering fine magnetic
CC particles formed in the magnetic microbe AMB-1. The magnetic microbe
CC can be used for the stable prepn. of proteins.
SQ Sequence 434 AA;

Query Match 54.8%; Score 57; DB 20; Length 434;
Best Local Similarity 55.6%; Pred. No. 7.62e+01;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 80 rflevwhta 88
QY :||| ||| :
1 KFLAVWKIT 9

Search completed: Wed Aug 4 15:30:16 1999
Job time : 14 secs.
```


SEQUENCE CHARACTERISTICS:

CC LENGTH: 13
CC TYPE: AMINO ACID
CC STRANDEDNESS: UNKNOWN
CC TOPOLOGY: UNKNOWN
CC MOLECULE TYPE: PEPTIDE
SQ SEQUENCE 13 AA; 1613 MW; 1196 CN;

Query Match 83.7%; Score 87; DB 2; Length 13;
Best Local Similarity 84.6%; Pred. No. 3.23e-02;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 KLFAYWKITYKDT 13
|::|::|::|::|::|::|::|
QY 1 KFLAVWKITYKDT 13

RESULT 2

ID US-08-479-400-5 STANDARD; PRT; 13 AA.

AC xxxxxx

Sequence 5, Application US/08479400

Sequence 5, Application US/08479400
Patent No. 5679784

GENERAL INFORMATION:

CC APPLICANT: Ladant, Daniel
CC APPLICANT: Leclerc, Claude
CC APPLICANT: Sebo, Peter
CC APPLICANT: Ullmann, Agnes
CC TITLE OF INVENTION: Recombinant Mutants for Inducing
CC TITLE OF INVENTION: Specific Immune Responses
CC NUMBER OF SEQUENCES: 7
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/479,400
CC FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/011,644
CC FILING DATE: 29-JAN-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 03495-0109-01000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 13 AA; 1613 MW; 1196 CN;

Query Match

• Best Local Similarity 83.7%; Score 87; DB 1; Length 13;
Best Local Similarity 84.6%; Pred. No. 3.23e-02;

Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 KLFAYWKITYKDT 13
|::|::|::|::|::|::|::|
QY 1 KFLAVWKITYKDT 13

RESULT 3

ID US-08-336-087-5 STANDARD; PRT; 13 AA.

AC xxxxxx

Sequence 5, Application US/08336087

Sequence 5, Application US/08336087
Patent No. 5503829

GENERAL INFORMATION:

CC APPLICANT: Ladant, Daniel
CC APPLICANT: Leclerc, Claude
CC APPLICANT: Sebo, Peter
CC APPLICANT: Ullmann, Agnes
CC TITLE OF INVENTION: Recombinant Mutants for Inducing
CC TITLE OF INVENTION: Specific Immune Responses
CC NUMBER OF SEQUENCES: 7
CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
CC ADDRESSEE: Dunner
CC STREET: 1300 I Street, N.W.
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005-3315

COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/336,087
CC FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/011,644
CC FILING DATE: 29-JAN-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Meyers, Kenneth J.
CC REGISTRATION NUMBER: 25,146
CC REFERENCE/DOCKET NUMBER: 03495-0109-01000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 13 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 13 AA; 1613 MW; 1196 CN;

Query Match

Best Local Similarity 83.7%; Score 87; DB 1; Length 13;
Best Local Similarity 84.6%; Pred. No. 3.23e-02;

Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 KLFAYWKITYKDT 13
|::|::|::|::|::|::|::|
QY 1 KFLAVWKITYKDT 13

RESULT 4

ID US-07-852-260-2 STANDARD; PRT; 2206 AA.

XX

CC CLASSIFICATION: 435
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Cseri, Luann
 CC REGISTRATION NUMBER: 31,822
 CC REFERENCE/DOCKET NUMBER: AVIR-011/OOUS
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: 415-494-7622
 CC TELEFAX: 415-857-0663
 CC INFORMATION FOR SEQ ID NO: 25:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 642 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC IMMEDIATE SOURCE:
 CC CLONE: tol.21
 CC FEATURE:
 CC NAME/KEY: protein
 CC LOCATION: 1..642
 CC OTHER INFORMATION: /label= ULL150
 CC SEQUENCE 642 AA; 70848 MW; 2000477 CN;

Query Match 58.7%; Score 61; DB 1; Length 642;
 Best Local Similarity 63.6%; Pred. No. 1.58e+01;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 404 LAIWQRTYNDI 414
 ||:| ||:|
 QY 3 LAVWKITYKDT 13

RESULT 7
 ID US-08-785-071A-2 STANDARD; PRT; 876 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 2, Application US/08785071A
 XX Sequence 2, Application US/08785071A
 CC Patent No. 5776750
 CC GENERAL INFORMATION:
 CC APPLICANT: Hodgson, John
 CC APPLICANT: Lawlor, Elizabeth
 CC TITLE OF INVENTION: No. 5776750el tRNA Synthetase
 CC NUMBER OF SEQUENCES: 2
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: SmithKline Beecham Corporation
 CC STREET: 709 Swedeland Road
 CC CITY: King of Prussia
 CC STATE: PA
 CC COUNTRY: USA
 CC ZIP: 19406-0939
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM Compatible
 CC OPERATING SYSTEM: DOS
 CC SOFTWARE: FastSeq for Windows Version 2.0
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/785,071A
 CC FILING DATE: 17-JAN-1997
 CC CLASSIFICATION: 424
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: 9601099.6
 CC FILING DATE: 19-JAN-1996
 CC APPLICATION NUMBER: 9622617.0
 CC FILING DATE: 27-JUL-1996
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Gimmi, Edward R
 CC REGISTRATION NUMBER: 38,891
 CC REFERENCE/DOCKET NUMBER: P31355-6
 CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 610-270-4478
 CC TELEFAX: 610-270-5090
 CC TELEX:
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 876 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: peptide
 CC SEQUENCE 876 AA; 98538 MW; 3870699 CN;
 Query Match 55.8%; Score 58; DB 2; Length 876;
 Best Local Similarity 25.0%; Pred. No. 3.11e+01;
 Matches 3; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 197 RYLEVWNLVFSE 208
 :|| ||:| :
 QY 1 KFLAVWKITYKD 12

RESULT 8
 ID US-08-529-600D-2 STANDARD; PRT; 434 AA.
 XX
 AC xxxxxx
 XX
 DT
 XX
 DE Sequence 2, Application US/08529600D
 XX Sequence 2, Application US/08529600D
 CC Patent No. 5861285
 CC GENERAL INFORMATION:
 CC APPLICANT: Tadashi MATSUNAGA
 CC TITLE OF INVENTION: PROTEIN-BOUND MAGNETIC PARTICLES AND
 CC TITLE OF INVENTION: PROCESS OF PRODUCING THE SAME
 CC NUMBER OF SEQUENCES: 3
 CC CORRESPONDENCE ADDRESS:
 CC ADDRESSEE: PILLSBURY MADISON & SUTRO, L.L.P.
 CC STREET: 1100 NEW YORK AVENUE, N.W.
 CC CITY: WASHINGTON
 CC STATE: D.C.
 CC COUNTRY: U.S.A.
 CC ZIP: 20005-3918
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Diskette
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Microsoft Word
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/529,600D
 CC FILING DATE: 18-SEP-1995
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: JP 6-248700
 CC FILING DATE: 15-SEP-1994
 CC INFORMATION FOR SEQ ID NO: 2:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 434 amino acids
 CC TYPE: amino acid
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC SEQUENCE 434 AA; 46827 MW; 982549 CN;

Query Match 54.8%; Score 57; DB 2; Length 434;
 Best Local Similarity 55.6%; Pred. No. 3.89e+01;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Db 80 RYLEVWKTA 88
 :|| ||:| :
 QY 1 KFLAVWKIT 9

```

RESULT 9
ID US-08-470-298B-11 STANDARD; PRI; 132 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 11, Application US/08470298B
XX
CC Sequence 11, Application US/08470298B
CC Patent No. 5844081
CC GENERAL INFORMATION:
CC APPLICANT: NI, JIAN
CC APPLICANT: GENTZ, REINER
CC APPLICANT: YU, GUO-LIANG
CC APPLICANT: ROSEN, CRAIG A.
CC TITLE OF INVENTION: CYTOSTATIN I
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: HUMAN GENOME SCIENCES, INC.
CC STREET: 9410 KEY WEST AVENUE
CC CITY: ROCKVILLE
CC STATE: MD
CC COUNTRY: US
CC ZIP: 20850
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/470,298B
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BROOKES, ALLAN A.
CC REGISTRATION NUMBER: 36,373
CC REFERENCE/DOCKET NUMBER: PF175D1
CC TELEPHONE: 301-309-8504
CC TELEFAX: 301-309-8512
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 132 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC ORGANISM: MYELIN P2 (FIGURE 2)
CC SEQUENCE 132 AA; 14909 MW; 94859 CN;

Query Match 51.9%; Score 54; DB 2; Length 132;
Best Local Similarity 62.5%; Pred. No. 7.55e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 4 KFLGTWKL 11
QY 1 KFLAVWKI 8

RESULT 10
ID US-08-409-731A-11 STANDARD; PRI; 132 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 11, Application US/08409731A
XX
CC Sequence 11, Application US/08409731A
CC Patent No. 5658758
CC GENERAL INFORMATION:

```

```

CC APPLICANT: Ni, Jian
CC APPLICANT: Yu, Guo-Liang
CC APPLICANT: Gentz, Reiner
CC APPLICANT: Rosen, Craig A.
CC TITLE OF INVENTION: CYTOSTATIN I
CC NUMBER OF SEQUENCES: 11
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: HUMAN GENOME SCIENCES, INC.
CC STREET: 9410 KEY WEST AVENUE
CC CITY: ROCKVILLE
CC STATE: MD
CC COUNTRY: USA
CC ZIP: 20850
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/409,731A
CC FILING DATE: 24-MAR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Benson, Robert H
CC REGISTRATION NUMBER: 30,446
CC REFERENCE/DOCKET NUMBER: PF175
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 301-309-8504
CC TELEFAX: 301-309-8512
CC INFORMATION FOR SEQ ID NO: 11:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 132 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 132 AA; 14909 MW; 94859 CN;

Query Match 51.9%; Score 54; DB 1; Length 132;
Best Local Similarity 62.5%; Pred. No. 7.55e+01;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 4 KFLGTWKL 11
QY 1 KFLAVWKI 8

RESULT 11
ID US-08-241-664B-14 STANDARD; PRI; 25 AA.
XX
AC xxxxxx
XX
DT
DE
DE Sequence 14, Application US/08241664B
XX
CC Sequence 14, Application US/08241664B
CC Patent No. 5871909
CC GENERAL INFORMATION:
CC APPLICANT: Voorhees, John J.
CC APPLICANT: Astrom, Anders
CC APPLICANT: Pattersson, Ulrika
CC APPLICANT: Tavakkol, Amir
CC TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
CC STREET: PO Box 828
CC CITY: Bloomfield Hills
CC STATE: Michigan
CC COUNTRY: United States of America
CC ZIP: 48303
CC COMPUTER READABLE FORM:

```



```
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/241,664B
CC FILING DATE: May 11, 1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36,683
CC REFERENCE/DOCKET NUMBER: 2115-00676COD
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810) 641-1600
CC TELEFAX: (810) 641-0270
CC TELEX: 287637
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA; 2950 MW; 3278 CN;

Query Match 51.0%; Score 53; DB 2; Length 25;
Best Local Similarity 41.7%; Pred. No. 9.40e+01;
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 2 NFLENWKIIXSE 13
:| | | | :
QY 1 KFLAVWKITYKD 12

RESULT 12
ID US-08-468-709B-14 STANDARD; PRT; 25 AA.
XX xxxxxx
XX
DI
XX
DE Sequence 14, Application US/08458709B
XX
CC Sequence 14, Application US/08458709B
CC Patent No. 5654137
CC GENERAL INFORMATION:
CC APPLICANT: Astrom, Anders
CC APPLICANT: Voorhees, John
CC APPLICANT: Pattersson, Ulrika
CC APPLICANT: Tavakkol, Amir
CC TITLE OF INVENTION: HUMAN CRABP-I AND CRABP-II
CC NUMBER OF SEQUENCES: 18
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
CC STREET: PO Box 828
CC CITY: Bloomfield Hills
CC STATE: Michigan
CC COUNTRY: United States of America
CC ZIP: 48303
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,709B
CC FILING DATE: 06/06/95
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Smith, Deann F.
CC REGISTRATION NUMBER: 36,683
CC REFERENCE/DOCKET NUMBER: 2115-00676DVF
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (810) 641-1600
CC
```

```
CC TELEFAX: (810) 641-0270
CC TELEX: 287637
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 25 amino acids
CC TYPE: amino acid
CC TOPOLOGY: unknown
CC MOLECULE TYPE: peptide
CC SEQUENCE 25 AA; 2950 MW; 3278 CN;

Query Match 51.0%; Score 53; DB 1; Length 25;
Best Local Similarity 41.7%; Pred. No. 9.40e+01;
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 2 NFLENWKIIXSE 13
:| | | | :
QY 1 KFLAVWKITYKD 12

RESULT 13
ID US-07-879-617A-9 STANDARD; PRT; 635 AA.
XX xxxxxx
AC
XX
DT
XX
DE Sequence 9, Application US/07879617A
XX
CC Sequence 9, Application US/07879617A
CC Patent No. 5580775
CC GENERAL INFORMATION:
CC APPLICANT: Fremeau Jr., Robert T.
CC APPLICANT: Caron, Marc G.
CC APPLICANT: Blakely, Randy D.
CC TITLE OF INVENTION: A High Affinity L-Proline Transporter
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Kilpatrick & Cody
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: U.S.
CC ZIP: 30309
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/879,617A
CC FILING DATE: 19920501
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: EMU109
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-815-6508
CC TELEFAX: 404-815-6555
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 635 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Rattus
CC TISSUE TYPE: Brain - Proline Transporter
CC FEATURE:
```

CC NAME/KEY: Domain
CC LOCATION: 46..65
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 72..97
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 117..137
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 217..236
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 243..264
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 291..311
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 322..345
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 373..397
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 424..443
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 456..479
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 500..519
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 536..559
CC OTHER INFORMATION: /note= "Proposed transmembrane
CC OTHER INFORMATION: domain."
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 76..77
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 83..84
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site
CC LOCATION: 90..91
CC OTHER INFORMATION: /note= "Leucine zipper motif"
CC FEATURE:
CC NAME/KEY: Active-site

CC LOCATION: 97..98
CC OTHER INFORMATION: /note= "Leucine zipper motif"
SQ SEQUENCE 635 AA; 70857 MW; 2226371 CN;
Query Match 51.0%; Score 53; DB 1; Length 635;
Best Local Similarity 85.7%; Pred. No. 9.40e+01;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Db 106 LAVWKIS 112
|||
QY 3 LAVWKIT 9
RESULT 14
ID US-08-753-985-9 STANDARD; PRT; 635 AA.
XX xxxxxx
AC
XX
DT
XX
DE
XX
CC Sequence 9, Application US/08753985
CC Sequence 9, Application US/08753985
CC Patent No. 5759788
CC GENERAL INFORMATION:
CC APPLICANT: Fremieu Jr., Robert I.
CC APPLICANT: Caron, Marc G.
CC APPLICANT: Blakely, Randy D.
CC TITLE OF INVENTION: A High Affinity L-Proline Transporter
CC NUMBER OF SEQUENCES: 13
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Kilpatrick & Cody
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: U.S.
CC ZIP: 30309
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/753,985
CC FILING DATE: 03-DEC-1996
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/879617
CC FILING DATE: 01-MAY-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC REFERENCE/DOCKET NUMBER: EMU109
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-815-6508
CC TELEFAX: 404-815-6555
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 635 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHEICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Rattus
CC TISSUE TYPE: Brain - Proline Transporter
CC FEATURE:
CC NAME/KEY: Domain
CC LOCATION: 46..65
CC OTHER INFORMATION: /note= "Proposed transmembrane

This Page Blank (uspto)